

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:30:40 ; Search time 9.67742 Seconds
(without alignments)
688.462 Million cell updates/sec

Title: US-09-863-901-22
Perfect score: 258
Sequence: 1 SGGSGGGGSGGSGGSGG.....SGSGGGGSGGSGGSLRS 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID22/cgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/cgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/cgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/cgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/cgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/cgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/cgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/cgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/cgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/cgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/cgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/cgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/cgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/cgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/cgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/cgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/cgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/cgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/cgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/cgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/cgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/cgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/cgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	50	ABB08752	GFP cloning interm
2	258	100.0	50	ABB08757	GFP peptide linker
3	258	100.0	477	ABB08634	GFP fusion protein
4	258	100.0	906	ABB08635	GFP fusion protein
5	212	82.2	468	ABB08633	GFP fusion protein
6	211	81.8	41	ABB08756	GFP peptide linker
7	170	65.9	738	AAW56163	New DNA sequence i
8	154	59.7	235	AAW05704	Glycine-rich repea
9	154	59.7	235	AAW79126	Epstein Barr Virus
10	154	59.7	261	AAW79137	FLGA Gly-ala inser

11	154	59.7	641	20	AAW28843	Epstein Barr Virus
12	154	59.7	641	21	AAW5856	Epstein Barr virus
13	154	59.7	641	22	AAW62332	EBV tethering prot
14	150	58.1	60	22	AAW77867	MHC class II H dom
15	150	58.1	60	22	AAW05183	Pain-relieving tar
16	148	57.4	265	19	AAW79128	FLGA insert stabil
17	147.5	57.2	532	23	AAW69423	Lung small cell ca
18	147.5	57.2	533	22	AAW05812	Human small cell l
19	147	57.0	56	23	ABP29010	Streptococcus poly
20	147	57.0	357	22	ABB29912	Peptide #2563 enco
21	147	57.0	357	22	ABB35090	Peptide #2596 enco
22	147	57.0	357	22	ABB20509	Protein #2508 enco
23	147	57.0	357	22	AAW55912	Human brain. expres
24	147	57.0	357	22	AAW68282	Human bone marrow
25	147	57.0	357	22	AAW16105	Peptide #2539 enco
26	147	57.0	357	22	AAW28596	Peptide #2633 enco
27	147	57.0	357	22	AAW03832	Peptide #2514 enco
28	147	57.0	357	23	ABG37823	Human peptide enco
29	146	56.6	399	22	ABG28641	Novel human diago
30	141.5	54.8	312	18	AAW18564	Novel fusion prote
31	140	54.3	40	21	AAW87573	Linker between CH3
32	140	54.3	610	22	ABW67896	Drosophila melanog
33	139.5	54.1	618	21	AAW56803	Human prostate can
34	139	53.9	358	21	AAW50939	Human adult skin c
35	139	53.9	369	21	AAW50940	Human adult skin c
36	139	53.9	387	21	AAW38324	Human secreted pro
37	139	53.9	479	21	AAW50941	Human adult skin c
38	137	53.1	283	22	ABB30839	Peptide #3490 enco
39	137	53.1	283	22	ABB36017	Peptide #3523 enco
40	137	53.1	283	22	ABB21422	Protein #3421 enco
41	137	53.1	283	22	AAW56810	Human brain expres
42	137	53.1	283	22	AAW69190	Human bone marrow
43	137	53.1	283	22	AAW17023	Peptide #3457 enco
44	137	53.1	283	22	AAW29514	Peptide #3551 enco
45	137	53.1	283	22	AAW04731	Peptide #3413 enco

ALIGNMENTS

RESULT 1

ABB08752
ID ABB08752 standard; Protein; 50 AA.

AC ABB08752;

XX DT 03-MAY-2002 (first entry)

XX DT GFP cloning intermediate protein sequence 9.

XX DE GFP; aquorin; green fluorescent protein; photoprotein; bioluminescence;
KW Chemiluminescence Resonance Energy Transfer; CRET;
KW central nervous system; neural network.

XX OS Synthetic.

XX PH Key Location/Qualifiers

XX FT Region 1..9 /note= "9 amino acid repeat region"

XX FT Misc-difference 1..45 /note= "region optionally deleted for 1-4 copies of the 9 amino acid repeats"

XX FT WO200192300-A2.

XX PD 06-DEC-2001.

XX PF 01-JUN-2001; 2001WO-EP07057.

XX PR 01-JUN-2000; 2000US-208314P.

XX PR 09-JUN-2000; 2000US-210526P.

XX PR 14-DEC-2000; 2000US-255111P.

PR 09-JUN-2000; 2000US-210526P.
 PR 14-DEC-2000; 2000US-255111P.
 XX
 XX (INSP) INST PASTEUR.
 PA (CNRS) CENT NAT RECH SCI.
 XX
 XX Baubet V, Le Mouelllic H, Brulet P;
 PI
 XX WPI; 2002-139605/18.
 DR N-PSDB; ABA97888.
 DR
 XX
 PT New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells
 XX
 XX Claim 5; Page 31; 6lpp; English.
 PS
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEQ
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network.
 XX
 SQ Sequence 468 AA;
 Query Match 82.2%; Score 212; DB 23; Length 468;
 Best Local Similarity 97.6%; Pred. No. 3.5e-12;
 Matches 41; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 9 QSGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50
 :|||||
 DB 238 KSGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 279
 RESULT 6
 ABB08756
 ID ABB08756 standard; peptide; 41 AA.
 XX
 XX ABB08756;
 AC
 XX
 XX 03-MAY-2002 (first entry)
 DT
 XX
 DE GFP peptide linker SEQ ID NO 21.
 XX
 XX GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
 KW Chemiluminescence Resonance Energy Transfer; CRET;
 KW central nervous system; neural network.
 XX
 OS Synthetic.
 XX
 XX WO200192300-A2.
 PN
 XX
 XX 06-DEC-2001.
 PD
 XX
 XX 01-JUN-2001; 2001WO-EP07057.
 PF
 XX
 PR 01-JUN-2000; 2000US-208314P.
 PR 09-JUN-2000; 2000US-210526P.
 PR 14-DEC-2000; 2000US-255111P.
 XX
 XX (INSP) INST PASTEUR.
 PA

PA (CNRS) CENT NAT RECH SCI.
 XX
 XX Baubet V, Le Mouelllic H, Brulet P;
 DR
 XX WPI; 2002-139605/18.
 XX
 PT New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells
 XX
 XX Claim 23; Page 35; 6lpp; English.
 PS
 XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEQ
 CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network. The
 CC present sequence is that of a GFP peptide linker sequence of the
 CC invention.
 XX
 SQ Sequence 41 AA;
 Query Match 81.8%; Score 211; DB 23; Length 41;
 Best Local Similarity 100.0%; Pred. No. 5.6e-13;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 50
 :|||||
 DB 1 SGGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRS 41
 RESULT 7
 AAW56163
 ID AAW56163 standard; Protein; 738 AA.
 XX
 XX AAW56163;
 AC
 XX
 XX 28-JUL-1998 (first entry)
 DT
 XX
 DE New DNA sequence isolated from Pinctada fucata.
 XX
 XX Pinctada fucata; protein; cosmetic.
 KW
 XX
 OS Pinctada fucata.
 XX
 XX JPI0080285-A.
 PN
 XX
 XX 31-MAR-1998.
 PD
 XX
 XX 28-MAY-1997; 97JP-0138461.
 PF
 XX
 XX 15-JUL-1996; 96JP-0184459.
 PR
 XX (MIKI-) MIKIMOTO SEIYAKU KK.
 PA
 XX
 DR WPI; 1998-254410/23.
 DR N-PSDB; AAW22683.
 XX
 XX New cDNA and e.g. vector, host cell and polypeptide - used to
 PT produce polypeptide in high yields, which is used in cosmetics
 XX

RESULT 9

```

US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Lewis, Cheryl Y
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P

```

; INFORMATION FOR SEQ ID NO: 4:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 907 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-010-928B-4

Query Watch
Best Local Similarity 52.1%; Score 134.5; DB 2; Length 907;
Matches 27; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

Qy 2 GSGSGGQS-GSGSGGQSGGSGGQSGGSGGSGGSGGSGGSG 47
|||||:|
Db 611 GSGGAGGTCPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 657
|||||:|

```

;;
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center/Patent Division

RESULT 13
US-08-819-539-7
; Sequence 7, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,539
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-819-539-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;
Best Local Similarity 40.6%; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;
QY 3 GSGSGGSGSGSGGQ-----SGSGSGGSGG-SGSGGSGSGSGG 43
Db 141 GNGVGGANGAGGCGGLAELQIEIQLAQLGGGAGAGGAGGAGGAGGAGGAG 200
QY 44 GQSG 47
Db 201 GAGG 204

RESULT 14
US-09-030-270A-7
; Sequence 7, Application US/09030270A
; Patent No. 5977060
; GENERAL INFORMATION:
; APPLICANT: Zitter, Thomas A.
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: INSECT CONTROL WITH A
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square

CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,270A
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,226
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1521
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-030-270A-7

Query Match 48.3%; Score 124.5; DB 2; Length 344;
Best Local Similarity 40.6%; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;
QY 3 GSGSGGSGSGSGGQ-----SGSGSGGSGG-SGSGGSGSGSGG 43
Db 141 GNGVGGANGAGGCGGLAELQIEIQLAQLGGGAGAGGAGGAGGAGGAGGAG 200
QY 44 GQSG 47
Db 201 GAGG 204

RESULT 15
US-08-984-207-7
; Sequence 7, Application US/08984207
; Patent No. 6235974
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dewen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,207
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 60/033,230
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-984-207-7

Query Match      48.38; Score 124.5; DB 4; Length 344;
Best Local Similarity 40.68; Pred. No. 0.00017;
Matches 26; Conservative 7; Mismatches 12; Indels 19; Gaps 2;

QY      3 GSGSGGSGSGSGSGG-----SGSGSGSGSGG-SGSGSGSGSGSGG 43
      |:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      141 GNGVGGANGAKGAGGGGLAEALQETQILQLGGGAGAGAGGGVGGAGGADGSGGAG 200

QY      44 GQSG 47
      | |
Db      201 GAGG 204

```

Search completed: June 18, 2003, 17:32:56
Job time : 4.41536 secs

Query Match	Score	Score	DB	Length
Best Local Similarity	56.6%	Score 145;	DB 9;	Length 200;
Matches	26;	Pred. No. 3.3e-05;		
Conservative	0;	Mismatches	20;	Indels
				Gaps

Search completed: June 18, 2003, 17:44:07
Job time : 7.07211 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:30:43 ; Search time 41.9355 Seconds
(without alignments)
768.720 Million cell updates/sec

Title: US-09-863-901-22

Perfect score: 258
Sequence: 1 SGGSGGGGSGSGGSGG.....SGSGGGGSGSGGSGGSLRS 50

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main.*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	50	22	US-09-863-901-22
2	258	100.0	477	22	US-09-863-901-5
3	258	100.0	906	22	US-09-863-901-6
4	235	91.1	45	22	US-09-863-901-25
5	212	82.2	468	22	US-09-863-901-4
6	211	81.8	41	22	US-09-863-901-21

7	170	65.9	300	21	US-09-708-427-19850	Sequence 19850, A
8	170	65.9	302	21	US-09-708-427-19849	Sequence 19849, A
9	168	65.1	592	27	US-60-360-039-17534	Sequence 17534, A
10	163	63.2	1307	21	US-09-708-427-4493	Sequence 4493, Ap
11	163	63.2	1349	21	US-09-708-427-4492	Sequence 4492, Ap
12	163	63.2	1422	21	US-09-708-427-4491	Sequence 4491, Ap
13	155.5	60.3	138	26	US-10-219-999-45027	Sequence 45027, A
14	154	59.7	235	9	US-08-529-190A-1	Sequence 1, Appl
15	154	59.7	235	13	US-08-970-900-1	Sequence 1, Appl
16	154	59.7	266	13	US-08-970-900-5	Sequence 5, Appl
17	154	59.7	267	11	US-08-733-369A-106	Sequence 106, App
18	154	59.7	603	3	US-07-865-030-1	Sequence 1, Appl
19	154	59.7	641	18	US-09-410-399-4	Sequence 4, Appl
20	154	59.7	641	21	US-09-791-537-91237	Sequence 91237, A
21	154	59.7	641	25	US-10-138-098-52	Sequence 52, Appl
22	153	59.3	1115	21	US-09-708-427-4736	Sequence 4736, Ap
23	153	59.3	1145	21	US-09-708-427-4735	Sequence 4735, Ap
24	153	59.3	1218	21	US-09-708-427-4734	Sequence 4734, Ap
25	150	58.1	50	18	US-09-421-971-73	Sequence 73, Ap
26	150	58.1	50	23	US-09-949-039-32	Sequence 32, Appl
27	150	58.1	54	18	US-09-421-971-64	Sequence 64, Appl
28	150	58.1	55	18	US-09-421-971-74	Sequence 74, Appl
29	150	58.1	59	1	PCT-US00-41324-13	Sequence 13, Appl
30	150	58.1	59	18	US-09-421-971-65	Sequence 65, Appl
31	150	58.1	60	18	US-09-421-971-75	Sequence 75, Appl
32	150	58.1	60	22	US-09-833-203-18	Sequence 18, Appl
33	150	58.1	100	1	PCT-US00-27794B-4	Sequence 4, Appl
34	148.5	57.6	171	21	US-09-411-067C-4	Sequence 4, Appl
35	147.5	57.2	532	22	US-09-708-427-6562	Sequence 6562, Ap
36	147.5	57.2	533	22	US-09-833-790-428	Sequence 428, App
37	147.5	57.2	533	21	US-09-489-101A-22	Sequence 22, Appl
38	147.5	57.2	533	21	US-09-791-537-142569	Sequence 142569, A
39	147	57.0	175	18	US-09-417-507-43604	Sequence 43604, A
40	147	57.0	188	18	US-09-417-507-22219	Sequence 22219, A
41	147	57.0	201	1	PCT-US01-00358-5	Sequence 5, Appl
42	147	57.0	201	1	PCT-US99-17885-11	Sequence 11, Appl
43	147	57.0	201	14	US-09-054-231-35	Sequence 35, Appl
44	147	57.0	201	15	US-09-130-858-11	Sequence 11, Appl
45	147	57.0	201	18	US-09-470-850-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-863-901-22
; Sequence 22, Application US/09863901
; GENERAL INFORMATION:
; APPLICANT: BAUBET, VALERIE
; APPLICANT: LE MOUILLIC, HERVE
; APPLICANT: BRULET, PHILIPPE
; TITLE OF INVENTION: CHIMERIC GFP-AEQUORIN AS BIOLUMINESCENT Ca++ REPORTERS
; FILE REFERENCE: 03495-0207-00000
; CURRENT APPLICATION NUMBER: US/09/863,901
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/208,314
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/210,526
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 60/255,111
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 50
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: sequence of linker
US-09-863-901-22

Db 668 SGGSGGQSGGSGGSGGQSGGSGGQSGGSGGQSGGLRS 717

```

; Sequence 25, Application US/09863901
; GENERAL INFORMATION:
; APPLICANT: BAUBET, VALERIE
; APPLICANT: LE MOUELLIC, HERVE
; APPLICANT: BRULET, PHILIPPE
; TITLE OF INVENTION: CHIMERIC GFP-AEQUORIN AS BIOLUMINESCENT Ca++ REPORTERS
; TITLE OF INVENTION: AT THE SINGLE CELL LEVEL
; FILE REFERENCE: 03495-0207-00000
; CURRENT APPLICATION NUMBER: US/09/863,901
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/208,314
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/210,526
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 60/255,111
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Linker
; OTHER INFORMATION: This sequence may encompass either 9, 18, 27, 36,
; OTHER INFORMATION: or 45 amino acids, with 9 or 45 amino acids being
; OTHER INFORMATION: more preferable
; US-09-863-901-25

```

```

; OTHER INFORMATION: more preferable
US-09-863-901-25

Query Match          91.1%; Score 235; DB 22; Length 45;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps

```

Best Local Similarity 100.0%; Pred. No. 2e-13; Length 23;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 2 GSGSGGQSGSGGSGGSGGSGGSGGSGGSGGSGGQ 46
Db 1 GSGSGGQSGSGGSGGSGGSGGSGGSGGSGGSGGQ 45

```

DB      1  GSGSGGQSGGSGGSGGSGGSGGSGGSGGSGGQSG 45

RESULT 5
US-09-863-901-4
; Sequence 4, Application US/09863901
; GENERAL INFORMATION:
; APPLICANT: BAUBET, VALERIE
; APPLICANT: LE MOUËLLIC, HERVE
; APPLICANT: BRULET, PHILIPPE
; TITLE OF INVENTION: CHIMERIC GFP-AEQUORIN AS BIOLUMINESCENT Ca++ REPORTERS
; TITLE OF INVENTION: AT THE SINGLE CELL LEVEL
; FILE REFERENCE: 03495-0207-00000
; CURRENT APPLICATION NUMBER: US/09/863,901
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/208,314
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/210,526
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 60/255,111
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-863-901-4

```

; ORGANISM: Aequorea victoria
US-09-863-901-4

Query Match	82.2%	Score 212;	DB 22;	Length 468;
-------------	-------	------------	--------	-------------

[illegible]


```

RESULT 12
K2C5_BOVIN STANDARD; PRT: 166 AA.
ID AC P04262;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II cytoskeletal 68 kDa, component IB (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=85128114; PubMed=6084625;
RA Jorcano J.L., Franz J.K., Franke W.W.;
RT "Amino acid sequence diversity between bovine epidermal cytokeratin
RT polypeptides of the basic (type II) subfamily as determined from cDNA
RT clones."
RL Differentiation 28:155-163(1984).
CC -I- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -I- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; K03534; AAA30601.1; -.
DR PIR; A02948; KRBO2B.
DR InterPro; IPRO01664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT NON_TER 1 1
FT DOMAIN <1 41 ROD.
FT DOMAIN 42 166 TAIL.
FT DOMAIN <1 41 COIL_2B.
SQ SEQUENCE 166 AA; 15505 MW; 4BFE495A7C6B4BE0 CRC64;
Query Match 55.2%; Score 142.5; DB 1; Length 166;
Best Local Similarity 66.7%; Pred. No. 9.6e-05;
Matches 36; Conservative 2; Mismatches 9; Indels 7; Gaps 3;
QY 2 GGSGGGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGL 48
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 87 GGSGFGGGSGIGSSGFGGGSGFGGGSGFGGGSGFGGGSGFGGGSGFGGGSGGSGGV 140
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
LORL_MOUSE STANDARD; PRT: 481 AA.
ID LORL_MOUSE
AC F18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lorlatin.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=90275605; PubMed=2190691;
RA Mehrlin T., Hohli D., Rothnagel J.A., Longley M.A., Bundman D.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:36:34 ; Search time 9.7723 Seconds
(without alignments)
1054.241 Million cell updates/sec

Title: US-09-863-901-22
Perfect score: 258
Sequence: 1 SGSGGGGGSGGGGGG.....SGSGGGGGSGGGGGSLRS 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL21:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	65.9	302	10 Q9SL09	Q9SL09 arabidopsis
2	170	65.9	738	5 Q02402	Q02402 pinctada fu
3	169	65.5	125	5 Q964C1	Q964C1 encephalito
4	168	65.1	592	16 Q9PF60	Q9PF60 xylella fas
5	166	64.3	102	5 Q964C0	Q964C0 encephalito
6	163	63.2	1422	10 Q9ZUR3	Q9ZUR3 arabidopsis
7	163	63.2	1745	10 Q9MAK1	Q9MAK1 arabidopsis
8	161.5	62.6	100	5 Q964C3	Q964C3 encephalito
9	161.5	62.6	117	5 Q964C2	Q964C2 encephalito
10	161.5	62.6	121	5 Q964C5	Q964C5 encephalito
11	161.5	62.6	138	5 Q964C4	Q964C4 encephalito
12	159.5	61.8	207	10 Q43522	Q43522 lycopersico
13	154	59.7	786	6 O18740	O18740 canis faml
14	153	59.3	1218	10 Q9XIH2	Q9XIH2 arabidopsis
15	150	58.1	251	3 Q9UVE7	Q9UVE7 yarrowia ll
16	149	57.8	157	5 Q9GQX8	Q9GQX8 plasmodium

17	148.5	57.6	171	10	004339	004339 arabidopsis
18	147.5	57.2	87	10	Q9LRI5	Q9LRI5 pisum sativ
19	146	56.6	291	10	Q39337	Q39337 brassica na
20	146	56.6	370	5	Q9BPK8	Q9BPK8 plasmodium
21	146	56.6	370	5	Q9BPK6	Q9BPK6 plasmodium
22	145.5	56.4	118	10	Q949R3	Q949R3 arabidopsis
23	145	56.2	277	5	Q9GQF0	Q9GQF0 plasmodium
24	145	56.2	302	5	Q9GQL9	Q9GQL9 plasmodium
25	145	56.2	302	5	Q9GQL8	Q9GQL8 plasmodium
26	145	56.2	317	5	Q9BH54	Q9BH54 plasmodium
27	144.5	56.0	221	10	065514	065514 arabidopsis
28	144	55.8	561	11	Q9CXH6	Q9CXH6 mus musculu
29	143.5	55.6	208	5	Q25949	Q25949 plasmodium
30	143.5	55.6	718	12	Q91TR1	Q91TR1 tupaiia herp
31	143	55.4	588	12	Q9QEK6	Q9QEK6 cynomolgus
32	142	55.0	486	11	Q8R0T9	Q8R0T9 mus musculu
33	142	55.0	698	12	Q8QXK8	Q8QXK8 ectocarpus
34	141.5	54.8	619	12	Q9IFQ9	Q9IFQ9 cynomolgus
35	141	54.7	139	5	Q9GQX9	Q9GQX9 plasmodium
36	141	54.7	208	5	Q9U0A0	Q9U0A0 plasmodium
37	141	54.7	396	10	065450	065450 arabidopsis
38	140	54.3	610	5	Q9V5V8	Q9V5V8 drosophila
39	139.5	54.1	637	11	Q9D2K8	Q9D2K8 mus musculu
40	139	53.9	185	10	Q948R3	Q948R3 oryza sativ
41	138.5	53.7	271	10	Q08529	Q08529 nicotiana t
42	138.5	53.7	588	12	Q9IPQ8	Q9IPQ8 cynomolgus
43	138	53.5	164	5	Q9BJO5	Q9BJO5 plasmodium
44	138	53.5	484	16	053394	053394 mycobacteri
45	137.5	53.3	175	10	Q9LSN6	Q9LSN6 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9SL09	PRELIMINARY;	PRT;	302 AA.
AC	Q9SL09;			
DT	01-MAY-2000 (TREMREL. 13, Created)			
DT	01-MAY-2000 (TREMREL. 13, Last sequence update)			
DT	01-DEC-2001 (TREMREL. 19, Last annotation update)			
DE	AT2G05580 protein.			
GN	AT2G05580.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RX	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,			
RA	Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,			
RA	Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,			
RA	Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,			
RA	Salzberg S.L., Fraser C.M., Venter J.C.;			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RL	thaliana.";			
RL	Nature 402:761-768(1999).			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Lin X.;			
DR	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC006220; AAD24649.1; -.			
DR	InterPro; IPR002952; Eggshell.			
DR	PRINTS; PR01228; EGGSHLL.			
DR	PRINTS; PR00341; PRON			
SQ	SEQUENCE 302 AA; 26008 MW; E72A73C55825891E CRC64;			

Search completed: June 17, 2003, 16:43:27
Job time : 12.7723 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:30:40 ; Search time 92.3226 Seconds
(without alignments)
688.462 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGBELFTGVPLVLDG.....LGFWTMDPACELKYGAVP 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	100.0	477	23	AB08634
2	2560	100.0	906	23	GFP fusion protein
3	2498.5	97.6	468	23	GFP fusion protein
4	2395.5	93.6	450	23	GFP fusion protein
5	2344	91.6	441	23	GFP fusion protein
6	2287.5	89.4	432	23	GFP fusion protein
7	1353	52.9	655	23	Protein related to
8	1295	50.6	719	19	Smad2-green floure
9	1295	50.6	719	21	EGFP-SMAD2 fusion
10	1293.5	50.5	356	23	yeast polypeptide

11	1292.5	50.5	631	19	AAW85006
12	1292.5	50.5	631	21	AAW70778
13	1291.5	50.4	1386	21	AAW85573
14	1290.5	50.4	890	21	AAW22938
15	1288.5	50.3	294	21	AAW22860
16	1288.5	50.3	294	21	AAW79638
17	1286	50.2	1070	18	AAW17789
18	1283.5	50.1	633	19	AAW85009
19	1283	50.1	459	21	AAW22936
20	1278.5	49.9	635	21	AAW85034
21	1278.5	49.9	635	21	AAW70781
22	1277	49.9	501	18	AAW31879
23	1276.5	49.9	1171	19	AAW85037
24	1274	49.8	250	22	AAW86492
25	1272	49.7	558	19	AAW48662
26	1271.5	49.7	1090	21	AAW85576
27	1271	49.6	397	20	AAW42176
28	1271	49.6	429	20	AAW42175
29	1271	49.6	501	18	AAW31877
30	1271	49.6	501	18	AAW31878
31	1271	49.6	514	18	AAW31876
32	1270	49.6	238	18	AAW22101
33	1270	49.6	238	20	AAW96328
34	1270	49.6	247	20	AAW96329
35	1269.5	49.6	359	20	AAW42179
36	1269.5	49.6	359	20	AAW42180
37	1269.5	49.6	391	20	AAW42178
38	1269.5	49.6	997	19	AAW85032
39	1269	49.6	607	19	AAW85010
40	1269	49.6	783	21	AAW22939
41	1268	49.5	248	23	AAW86319
42	1268	49.5	403	20	AAW42177
43	1268	49.5	842	19	AAW85008
44	1267	49.5	605	19	AAW85007
45	1267	49.5	727	19	AAW85041

ALIGNMENTS

RESULT 1

AB08634

ID AB08634 standard; Protein; 477 AA.

XX AC AB08634;

XX DT 03-MAY-2002 (first entry)

XX DE GFP fusion protein G5A SEQ ID NO 5.

XX KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;

XX KW Chemiluminescence Resonance Energy Transfer; CRET;

XX KW central nervous system; neural network.

XX OS Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 445

XX FT /note= "Encoded by AQT"

XX PN WO200192300-A2.

XX PD 06-DEC-2001.

XX PF 01-JUN-2001; 2001WO-EP07057.

XX PF 01-JUN-2000; 2000US-208314P.

XX PR 09-JUN-2000; 2000US-210526P.

XX PR 14-DEC-2000; 2000US-255111P.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CENT NAT RECH SCI.

Erk1-green floures
EGFP-Erk1 fusion p
Hs-UNC-53/3 fragm
GFP-HSC70 fusion p
GFP-DEV-annexin I
Caspase-3 biosenso
Green fluorescent
Jnk1-green floures
GFP-HSP27 fusion p
Green fluorescent
EGFP-VASP fusion p
GFP variants S65T
Green fluorescent
GR fusion protein
Hs-UNC-53/1 fragm
EGFP/DRM fusion pr
EGFP/DRM fusion pr
GFP variants S65C
GFP variants S65C
Aequorea victoria
Humanised green fl
EGFP/DRM fusion pr
EGFP/DRM fusion pr
EGFP/DRM fusion pr
Green fluorescent
p38-green flouresc
GFP-HSP1 fusion pr
Jellyfish green fl
EGFP/DRM fusion pr
GK5-green floures
Erk2-green floures
Green fluorescent

PI Baubet V, Le Mouelllic H, Brulet P;
 XX WPI; 2002-139605/18.
 DR N-PSDB; ABA97889.
 XX
 XX New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells -
 XX
 XX
 PS Claim 6; Page 31; 61pp; English.
 XX
 CC The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEO
 CC GFP = green fluorescent protein; AEO = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network.
 XX
 XX Sequence 477 AA;
 SQ
 Query Match 100.0%; Score 2560; DB 23; Length 477;
 Best Local Similarity 100.0%; Pred. No. 1.4e-207;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFGVVPILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVWPPL 60
 DB 1 MSKGEELFGVVPILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVWPPL 60
 QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 DB 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
 QY 181 HYQONTPIGDPVLLPDNHYLSTQALSADPNKRDHMLVLEFVTAAGTTHGMDELYKSG 240
 DB 181 HYQONTPIGDPVLLPDNHYLSTQALSADPNKRDHMLVLEFVTAAGTTHGMDELYKSG 240
 QY 241 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 300
 DB 241 GSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 300
 QY 301 IGRHKHMFNFDVNNKISLDEMVKASDIIVNINLGATPEQAKRHKDAVEAFFGAGMK 360
 DB 301 IGRHKHMFNFDVNNKISLDEMVKASDIIVNINLGATPEQAKRHKDAVEAFFGAGMK 360
 QY 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFIDVDKQNGAITLDEWAKY 420
 DB 361 YGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGDALFIDVDKQNGAITLDEWAKY 420
 QY 421 TAAAGTIQSSDECEETFRVCDIDESGQLDVDEMTQHLGFWTMDPACBKLKYGAVP 477
 DB 421 TAAAGTIQSSDECEETFRVCDIDESGQLDVDEMTQHLGFWTMDPACBKLKYGAVP 477
 RESULT 2
 ID ABB08635 standard; Protein; 906 AA.
 XX ABB08635
 AC ABB08635;

XX 03-MAY-2002 (first entry)
 XX GFP fusion protein Seg5A SEQ ID NO 6.
 XX
 KW GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;
 KW Chemiluminescence Resonance Energy Transfer; CRET;
 KW central nervous system; neural network.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 305 /note= "Encoded by QGT"
 FT Misc-difference 623 /note= "Encoded by CQT"
 FT Misc-difference 868 /note= "Encoded by GTQ"
 FT
 XX WO2001923300-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 01-JUN-2001; 2001WO-EP07057.
 XX
 XX 01-JUN-2000; 2000US-208314P.
 PR 09-JUN-2000; 2000US-210526P.
 PR 14-DEC-2000; 2000US-255111P.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CENT NAT RECH SCI.
 XX
 XX Baubet V, Le Mouelllic H, Brulet P;
 XX WPI; 2002-139605/18.
 DR N-PSDB; ABA97889.
 XX
 XX New fusion protein comprising a modified bioluminescent system with a
 PT fluorescent molecule covalently linked with a photoprotein, useful for
 PT monitoring calcium fluxes or for detecting electrical activity in a
 PT group of neural cells -
 XX
 XX
 PS Claim 7; Page 31-32; 61pp; English.
 XX
 CC The invention relates to a fusion protein (ABB08630-ABB08635) for energy
 CC transfer from aequorin to green fluorescent protein by Chemiluminescence
 CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule
 CC covalently linked with a photoprotein. The fusion protein comprises the
 CC formula: GFP - LINKER - AEO
 CC GFP = green fluorescent protein; AEO = aequorin; and LINKER = a
 CC polypeptide of 4-63 amino acids.
 CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the
 CC single cell level. The fusion protein, composition or bioluminescent
 CC system is useful for monitoring calcium fluxes in real time. This is
 CC particularly useful for understanding the development, the plasticity and
 CC the functioning of the central nervous system. The fusion protein
 CC comprising the bioluminescent system is useful for detecting electrical
 CC activity in a group of neural cells, for making it possible to complete
 CC the phenotype study of mutants, or for observing the calcium activity in
 CC a population of connected cells, for example in a neural network.
 XX
 XX Sequence 906 AA;
 SQ
 Query Match 100.0%; Score 2560; DB 23; Length 906;
 Best Local Similarity 100.0%; Pred. No. 3.3e-207;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFGVVPILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVWPPL 60
 DB 430 MSKGEELFGVVPILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVWPPL 489
 QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120

490	VTTLTYGVQVCSRPVPHDKHQHDFPKSAMPEGYVQERTIFFPKDDGNTKTRAEVFEQDTLV	549
121	NRIELKGIDFKREDGNILGHKLEYNYNSHNYYIMADKOKNGIKANFKIRHNIEDGSVOLAD	180
550	NRIELKGIDFKREDGNILGHKLEYNYNSHNYYIMADKOKNGIKANFKIRHNIEDGSVOLAD	609
181	HYQNTPTGDCGPVLLPNDHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYKSG	240
610	HYQNTPTGDCGPVLLPNDHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYKSG	669
241	GSGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGG	300
670	GSGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGG	729
301	IGRHKHMFNFLDVNHNGKISLDEMYYKASDIVINNLTGATPEQAKRHKDAVEAFFGGAGMK	360
730	IGRHKHMFNFLDVNHNGKISLDEMYYKASDIVINNLTGATPEQAKRHKDAVEAFFGGAGMK	789
361	YGVETDWPAYTEGKKKLTATDELYAKNEPTLIRINGDALFDIVDKDNGAITLDEWKAY	420
790	YGVETDWPAYTEGKKKLTATDELYAKNEPTLIRINGDALFDIVDKDNGAITLDEWKAY	849
421	TKAAGIIQSSDECEETFRVCDIDESGOLDVDEMTROHLGFWYTMDPACEKLYGGAVP	477
850	TKAAGIIQSSDECEETFRVCDIDESGOLDVDEMTROHLGFWYTMDPACEKLYGGAVP	906
RESULT 3		
ABB08633		
ID	ABB08633 standard; Protein; 468 AA.	
XX	AC	ABB08633;
DT	03-MAY-2002 (first entry)	
XX	CFP fusion protein G4A SEQ ID NO 4.	
KW	GFP; aquorin; green fluorescent protein; photoprotein; bioluminescence;	
KW	Chemiluminescence Resonance Energy Transfer; CREW;	
KW	central nervous system; neural network.	
OS	Synthetic.	
XX	Key	Location/Qualifiers
FT	Misc-difference 401	
FT	/note= "Encoded by GGA"	
XX	WO200192300-A2.	
XX	06-DEC-2001.	
XX	01-JUN-2001; 2001WO-EP07057.	
PR	01-JUN-2000; 2000US-208314P.	
PR	09-JUN-2000; 2000US-210526P.	
PR	14-DEC-2000; 2000US-255111P.	
XX	(INSP) INST PASTEUR.	
PA	(CNRS) CENT NAT RECH SCI.	
XX	Baubet V, Le Mouellic H, Brulet P;	
PI	WPI; 2002-139605/18.	
XX	N-PSDB; ABA97888.	
XX	New fusion protein comprising a modified bioluminescent system with a	
PT	fluorescent molecule covalently linked with a photoprotein, useful for	
PT	monitoring calcium fluxes or for detecting electrical activity in a	
PT	group of neural cells	
XX	Claim 5; Page 31; 61pp; English.	
PS	The invention relates to a fusion protein (ABB08630-ABB08635) for energy	
CC		

PD 06-DEC-2001.

XX 01-JUN-2001; 2001WO-EP07057.

XX 01-JUN-2000; 2000US-208314P.

PR 09-JUN-2000; 2000US-210526P.

PR 14-DEC-2000; 2000US-255111P.

XX (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

XX Baubet V, Le Mouelllic H, Brulet P;

XX WPI; 2002-139605/18.

DR N-PSDB; ABA97887.

XX New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for

PT monitoring calcium fluxes or for detecting electrical activity in a

PT group of neural cells

XX Synthetic.

PS Claim 4; Page 31; 61pp; English.

XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy

CC transfer from aequorin to green fluorescent protein by Chemiluminescence

CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule

CC covalently linked with a photoprotein. The fusion protein comprises the

CC formula: GFP - LINKER - AEQ

CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a

CC polypeptide of 4-63 amino acids.

CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the

CC single cell level. The fusion protein, composition or bioluminescent

CC system is useful for monitoring calcium fluxes in real time. This is

CC particularly useful for understanding the development, the plasticity and

CC the functioning of the central nervous system. The fusion protein

CC comprising the bioluminescent system is useful for detecting electrical

CC activity in a group of neural cells, for making it possible to complete

CC the phenotype study of mutants, or for observing the calcium activity in

CC a population of connected cells, for example in a neural network.

XX Sequence 450 AA;

XX Query Match 93.6%; Score 2395.5; DB 23; Length 450;

XX Best Local Similarity 94.3%; Pred. No. 9.8e-194;

XX Matches 450; Conservative 0; Mismatches 0; Indels 27; Gaps 1;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVPWPTL 60

DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVPWPTL 60

QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKGEGDTLV 120

DB 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFKDDGNYKTRAEVKGEGDTLV 120

QY 121 NRIELKGIDFKEDGNLGHKLEYNNSHNYIMADQKNGIKANFKIRHNIEDGSGVQLAD 180

DB 121 NRIELKGIDFKEDGNLGHKLEYNNSHNYIMADQKNGIKANFKIRHNIEDGSGVQLAD 180

QY 181 HYQQNTPIGDGPVLLPDNHYLTSQALSNDPNEKRDHMLLEFVTAAGTTHGMDLYKSG 240

DB 181 HYQQNTPIGDGPVLLPDNHYLTSQALSNDPNEKRDHMLLEFVTAAGTTHGMDLYK-- 238

QY 241 GSGSGGQSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRSVKTLSDFDNP 300

DB 239 -----SGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSLRSVKTLSDFDNP 273

QY 301 IGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGNATPEQAKRHKDAVEAFFGAGMK 360

DB 274 IGRHKHMFNFDVNHNGKISLDEMYKASDIVNNLGNATPEQAKRHKDAVEAFFGAGMK 333

QY 361 YGVETDWPAYIBGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKONGAITLDEWKAY 420

DB 334 YGVETDWPAYIBGWKKLATDELEKYAKNEPTLIRIWGDALFDIVDKONGAITLDEWKAY 393

QY 421 TKAAGIIOSSECEETFRVCDIDESGQLDVEDMTQHGLFWYTMDPACEKLYGGAVP 477

DB 394 TKAAGIIOSSECEETFRVCDIDESGQLDVEDMTQHGLFWYTMDPACEKLYGGAVP 450

RESULT 5

ABB08631

ID ABB08631 standard; Protein; 441 AA.

XX ABB08631;

AC ABB08631;

XX 03-MAY-2002 (first entry)

DT GFP fusion protein G1A SEQ ID NO 2.

DE GFP fusion protein G1A SEQ ID NO 2.

XX GFP; aequorin; green fluorescent protein; photoprotein; bioluminescence;

KW Chemiluminescence Resonance Energy Transfer; CRET;

KW central nervous system; neural network.

XX Synthetic.

OS WO200192300-A2.

XX 06-DEC-2001.

PD 01-JUN-2001; 2001WO-EP07057.

PR 01-JUN-2000; 2000US-208314P.

PR 09-JUN-2000; 2000US-210526P.

PR 14-DEC-2000; 2000US-255111P.

XX (INSP) INST PASTEUR.

PA (CNRS) CENT NAT RECH SCI.

XX Baubet V, Le Mouelllic H, Brulet P;

XX WPI; 2002-139605/18.

DR N-PSDB; ABA97886.

XX New fusion protein comprising a modified bioluminescent system with a

PT fluorescent molecule covalently linked with a photoprotein, useful for

PT monitoring calcium fluxes or for detecting electrical activity in a

PT group of neural cells

XX Claim 3; Page 30-31; 61pp; English.

XX The invention relates to a fusion protein (ABB08630-ABB08635) for energy

CC transfer from aequorin to green fluorescent protein by Chemiluminescence

CC Resonance Energy Transfer (CRET), which comprises a fluorescent molecule

CC covalently linked with a photoprotein. The fusion protein comprises the

CC formula: GFP - LINKER - AEQ

CC GFP = green fluorescent protein; AEQ = aequorin; and LINKER = a

CC polypeptide of 4-63 amino acids.

CC The fusion protein is useful as a bioluminescent Ca²⁺ reporter at the

CC single cell level. The fusion protein, composition or bioluminescent

CC system is useful for monitoring calcium fluxes in real time. This is

CC particularly useful for understanding the development, the plasticity and

CC the functioning of the central nervous system. The fusion protein

CC comprising the bioluminescent system is useful for detecting electrical

CC activity in a group of neural cells, for making it possible to complete

CC the phenotype study of mutants, or for observing the calcium activity in

CC a population of connected cells, for example in a neural network.

XX Query Match 91.6%; Score 2344; DB 23; Length 441;

XX Best Local Similarity 92.5%; Pred. No. 2.1e-189;

XX Matches 441; Conservative 0; Mismatches 0; Indels 36; Gaps 1;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVPWPTL 60

DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYGKLTFLFICTTGKLPVPWPTL 60

XX OS Unidentified.
XX PN WO200240528-AL.
XX PD 23-MAY-2002.
XX PF 26-OCT-2001; 2001WO-CN01506.
XX PR 26-OCT-2000; 2000CN-0125859.
XX PA (SHAN-) SHANGHAI BIO-CHEM INST CHINESE ACAD SCI.
XX PI Lu C, Huang J, Zhao Y, Zhang F, Chen X;
XX WPI; 2002-427084/45.
XX
XX Method for producing non-natural silk by Bombyx mori with modifying
XX heavy and light chains of fibroin in natural silk by DNA recombination
XX technology and protein engineering, applicable in sericulture and
XX textile industry
XX
XX Claim 4; Page 40-42; 50pp; Chinese.
XX
XX This invention relates to a silk which constitutes recombinant Bombyx
XX mori silk fibroin obtained by modification or variation at the heavy
XX or light chains of the fibroin. The method is for producing non-natural
XX silk, which is applicable in sericulture and textile industry. Such
XX modified silk has improved performance. The present sequence is a
XX protein related to the invention.
XX
XX Sequence 655 AA;
XX
Query Match 52.9%; Score 1353; DB 23; Length 655;
Best Local Similarity 86.0%; Pred. No. 1.7e-105;
Matches 264; Conservative 4; Mismatches 19; Indels 20; Gaps 5;
QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLCFICTTGKLPVPWPTL 60
DB 6 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLCFICTTGKLPVPWPTL 65
QY 61 VTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 66 VTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 125
QY 121 NRLEKGIIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180
DB 126 NRLEKGIIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 185
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYK 238
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKGI 245
QY 239 --SGSGSG--GGSGSGG-----SGQSGSGSGSGSGSGSGGQ-----SGSGSGSG 281
DB 246 PADCRSGGLGGGPGGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 304
QY 282 GQSGLRs 288
DB 305 GTGGLGS 311
RESULT 8
ID AAW85012 standard; Protein; 719 AA.
XX AC AAW85012;
XX AC AAW85012;
XX DT 08-FEB-1999 (first entry)
XX Smad2-green fluorescent protein fusion product.
XX Human; Smad2 gene; fusion protein; green fluorescent protein; GFP;

KW intracellular signalling; chimera.
XX
XX Chimeric - Aequorea victoria.
XX OS Chimeric - Homo sapiens.
XX PN WO9845704-A2.
XX PD 15-OCT-1998.
XX PF 07-APR-1998; 98WO-DK00145.
XX PR 07-APR-1997; 97DK-0000392.
XX PA (NOVO) NOVO-NORDISK AS.
XX PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;
XX Tullin S;
XX WPI; 1998-594491/50.
XX DR N-PSDB; AAV71027.
XX
XX Determining effect on signalling pathways in live cells from
XX redistribution of luminophores - specifically fusions of green
XX fluorescent protein with a signalling component, and new apparatus,
XX particularly for identifying toxins and potential therapeutic agents
XX
XX Example 5; Pages 102-103; 326pp; English.
XX
XX The present sequence represents a human Smad2-green fluorescent
XX protein fusion product. The fusion protein is used in an assay
XX that exemplifies the invention. The specification describes how
XX quantitative information about the influence of a molecule on a cellular
XX response is obtained by recording the variation, caused by the molecule,
XX on mechanically intact living cells, in the spatially distributed light
XX emitted from a luminophore present in the cells. The variation in light
XX emission is processed to provide information that correlates spatial
XX distribution to the degree of the molecule. The method is used to
XX identify agents that (in)directly affect intracellular signalling,
XX especially to screen for potential therapeutic agents or toxins, and
XX to identify new drug targets.
XX
XX Sequence 719 AA;
Query Match 50.6%; Score 1295; DB 19; Length 719;
Best Local Similarity 81.0%; Pred. No. 1.5e-100;
Matches 251; Conservative 9; Mismatches 36; Indels 14; Gaps 1;
QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLCFICTTGKLPVPWPTL 60
DB 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYKLTLCFICTTGKLPVPWPTL 61
QY 61 VTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 62 VTTLTYGVQCSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRLEKGIIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180
DB 122 NRLEKGIIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 241
QY 241 GSGSGSGSGSGS-----GSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 286
DB 242 LNSRAQASNSTMSSILPPTPPVVKLLGWKKSSAGSGGGGQEQKCEKCAVKSIV 301
QY 287 RSVKLTSDFD 296
DB 302 KKLKKTGRLD 311
RESULT 9

AA70779
ID AAY70779 standard; Protein: 719 AA.
XX
AC AAY70779;
XX
DT 31-JUL-2000 (first entry)
XX
DE EGFP-SMAD2 fusion protein construct.
XX
KW Fusion construct; EGFP-SMAD2; Green fluorescent protein; cytokine;
XX cellular response; luminophore; screening; medicament; signal transducer;
KW light/fluorescence intensity; intracellular signalling pathway; mutant;
KW fluorescent probe; signal transduction; transforming growth factor-beta;
KW TGF-beta.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Aequorea victoria.
XX
PN WO200023615-A2.
XX
PD 27-APR-2000.
XX
PF 15-OCT-1999; 99WO-DK00562.
XX
PR 15-OCT-1998; 98DK-0001320.
XX
PA (BIOI-) BIOIMAGE AS.
XX
PI Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;
PI Hagel G;
XX
DR WPI: 2000-339710/29.
DR N-PSDB: RAD00037.
XX
PT Quantifying cellular response to an influence, useful for detecting
PT intracellular translocation or redistribution of biologically active
PT substances comprises recording changes in spatially distributed light
PT emitted from a luminophore
XX
PS Example 4; Page 110-112; 150pp; English.
XX
CC The patent discloses a method for extracting quantitative information
CC relating to an influence on a cellular response in mechanically intact
CC living cells, by recording variation in spatially distributed light
CC emitted from a luminophore, as a change in light/fluorescence intensity.
CC The luminophore present in the cells, is capable of being redistributed
CC in a manner which is related with the degree of the influence. This
CC method is useful as a screening program, for the identification of a
CC biologically active substance, that directly or indirectly affects an
CC intracellular signalling pathway. It is also potentially useful as a
CC medicament. The fluorescent probe is useful in the back-tracking of
CC signal transduction pathways. The present sequence is a fusion protein
CC construct EGFP-SMAD2, comprising the human Smad2 gene, fused to a
CC derivative of the luminescent green fluorescent protein, EGFP. Smad2 is
CC a signal transducer, that is induced by some members of the transforming
CC growth factor-beta (TGF) family of cytokines. This construct is under the
CC control of a CMV promoter and functions as the luminophore, that is
CC useful to monitor signalling pathways and identify compounds that
CC modulate the pathways in living cells.
XX
SQ Sequence 719 AA;

Query Match 50.68; Score 1295; DB 21; Length 719;
Best Local Similarity 81.08; Pred. No. 1.5e-100;
Matches 251; Conservative 9; Mismatches 16; Indels 14; Gaps 1;
QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGGDATYGLTKLTKFKICTTGKLPVWPPTL 60
DB 2 VSRGEELFTGVVPIVLVDGVDNGHFKFSVSGEGGDATYGLTKLTKFKICTTGKLPVWPPTL 61
QY 61 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 62 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

QY 121 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKOKNGIKANFKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLEFFVTAAGITTHGMDLYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTOSALSCKDPNEKRDHMLVLEFFVTAAGITTHGMDLYKSG 241
QY 241 GSG 286
DB 242 LRSRAQASNSMTSSILPTPPVVRLLGWKKSAGSGSGAGGEGQNGQEKWCEKAVKSILV 301
QY 287 RSVKLTSDFD 296
DB 302 KKLKKTGRLD 311

RESULT 10
ABB08620
ID ABB08620 standard; Protein: 356 AA.
XX
AC ABB08620;
XX
DT 10-APR-2002 (first entry)
XX
DE Yeast polypeptide 1.
XX
KW Yeast; drug; cell wall; GPI anchor protein.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200183733-A1.
XX
PD 08-NOV-2001.
XX
PF 26-APR-2001; 2001WO-JP03630.
XX
PR 01-MAY-2000; 2000JP-0132041.
XX
PA (DAUC) DAIICHI PHARM CO LTD.
XX
PI Kitamura A, Someya K, Nakajima R;
XX
DR WPI: 2002-097496/13.
XX
DR N-PSDB: ABA97833.
XX
PT Screening for drugs that act on cell walls, involves culturing
PT microorganisms with a reporter protein acting as a GPI anchor protein
PT in their cell walls -
XX
PS Disclosure; Fig 4-5; 44pp; Japanese.

The invention relates to screening for drugs that act on cell walls, comprising culturing microorganisms with a reporter protein acting as a GPI anchor protein in their cell walls in the presence of the test substance and assaying the amount of reporter protein produced in the culture. The present sequence is that of a polypeptide useful to the invention.

Query Match 50.5%; Score 1293.5; DB 23; Length 356;
Best Local Similarity 88.0%; Pred. No. 7.8e-101;
Matches 250; Conservative 3; Mismatches 26; Indels 5; Gaps 2;
QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGGDATYGLTKLTKFKICTTGKLPVWPPTL 60
DB 40 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGGDATYGLTKLTKFKICTTGKLPVWPPTL 99
QY 61 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 100 VTTLTGYVQCFSRYPDHMKQHDFFKFSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 159

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
 DB 160 NRIELKGIDFKEDGNILGHKLEYNNSHVYITADKQKNGIKANFKIRHNIEDGSVQLAD 219
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYK-- 238
 DB 220 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYK 279
 QY 239 --SGSGGG 280
 DB 280 LGSCTESSPASSNAGSSSKS-NSGSGSGSSSSSSSSSSSSSSSSSSSSSSSSSS 322

RESULT 11

AAW85006
 ID AAW85006 standard; Protein; 631 AA.

XX
 AC AAW85006;

XX
 DT 08-FEB-1999 (first entry)

XX
 DE Erkl-green fluorescent protein fusion product.

XX
 KW Human; Erkl gene; fusion protein; green fluorescent protein; GFP;

XX
 KW Intracellular signalling; chimera.

XX
 OS Chimeric - Aequorea victoria.

XX
 OS Chimeric - Homo sapiens.

XX
 PN W09845704-A2.

XX
 PD 15-OCT-1998.

XX
 PF 07-APR-1998; 98WO-DK00145.

XX
 PR 07-APR-1997; 97DK-0000392.

XX
 PA (NOVO) NOVO-NORDISK AS.

XX
 PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O;

XX
 PI Tullin S;

XX
 DR WPI: 1998-594491/50.

XX
 DR N-PSDB; AAW71022.

XX
 PT Determining effect on signalling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents

XX
 PS Example 3; Pages 69-71; 326pp; English.

XX
 CC The present sequence represents a human Erkl-green fluorescent
 CC protein fusion product. The fusion protein is used in an assay
 CC that exemplifies the invention. The specification describes how
 CC quantitative information about the influence of a molecule on a cellular
 CC response is obtained by recording the variation, caused by the molecule,
 CC on mechanically intact living cells, in the spatially distributed light
 CC emitted from a luminophore present in the cells. The variation in light
 CC emission is processed to provide information that correlates spatial
 CC distribution to the degree of the molecule. The method is used to
 CC identify agents that (indirectly) affect intracellular signalling,
 CC especially to screen for potential therapeutic agents or toxins, and
 CC to identify new drug targets.

XX
 SQ Sequence 631 AA;

Query Match 50.5%; Score 1292.5; DB 19; Length 631;
 Best Local Similarity 58.0%; Pred. No. 2.1e-100;
 Matches 283; Conservative 27; Mismatches 107; Indels 71; Gaps 11;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTFLKFICTTGKLPVWPPTL 60

DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATYKLTFLKFICTTGKLPVWPPTL 61
 QY 61 VTTLYGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNTKTRAEVKFEGDTLV 120
 DB 62 VTTLYGVQCFRYPDHMKQHDFFKSAMPEGVVOERTIFFKDDGNTKTRAEVKFEGDTLV 121
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
 DB 122 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 181
 QY 181 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYKSG 240
 DB 182 HYQONTPIGDPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGTTHGMDLYKSG 241
 QY 241 --SGSGGG 280
 DB 242 LRSRAQASNSTWAAAAAAGGGGGGPRTEGVPGEVEMVKGQDFVGPRTQLOIYIG 301
 QY 281 GGSGGLRS-----VKLTDFDNPRTWG-----RHKMFNFDLVNHN 316
 DB 302 EGAYGMVSSAYDHRVTRVAKKISPFHQYTCQRTLRITQILLRPHENNVIGIRDLRA 361
 QY 317 GKISLDEMVKASDIVINNLGATPEQAKRHKDAVEAFPG--AGMKYGVETDWPAYIEGW 374
 DB 362 STLEAMRDVYIVQDLMETDLYKLLKSQSLSDHICFYLOILRGLK-----YIHS 412
 QY 375 KKLATD-ELEKYAKNEPTLRIWGDALFIVD--KDONGAIT---LDEN-----K 418
 DB 413 NVLHRDLKPSNLLSNVTCDLKICDFGLARIADPEHDHGTGFLTEYVATRYRAPEIMLSK 472
 QY 419 AYTKAAGI 426
 DB 473 GYTKSIDI 480

RESULT 12

AAW70778
 ID AAW70778 standard; Protein; 631 AA.

XX
 AC AAW70778;

XX
 DT 31-JUL-2000 (first entry)

XX
 DE EGFP-Erkl fusion protein construct.

XX
 KW Fusion construct; EGFP-Erkl; Green fluorescent protein; luminophore;
 KW extracellular signal regulated kinase-1; serine/threonine protein kinase;
 KW mitogen activated protein kinase; MAPK; screening; medicament; mutant;
 KW cellular response; light/fluorescence intensity; signal transduction;
 KW intracellular signalling pathway; fluorescent probe.

XX
 OS Chimeric - Homo sapiens.

XX
 OS Chimeric - Aequorea victoria.

XX
 PN W0200023615-A2.

XX
 PD 27-APR-2000.

XX
 PF 15-OCT-1999; 99WO-DK00562.

XX
 PR 15-OCT-1998; 98DK-0001320.

XX
 PA (BIOI-) BIOIMAGE AS.

XX
 PI Arkhammar POG, Terry BR, Scudder KM, Bjorn SP, Thastrup O;

XX
 PI Hagel G;

XX
 DR WPI: 2000-339710/29.

XX
 DR N-PSDB; AAD00036.

PT Quantifying cellular response to an influence, useful for detecting
 PT intracellular translocation or redistribution of biologically active

PN XX WO200050872-A2.

PD XX 31-AUG-2000.

PF XX 25-FEB-2000; 2000WO-US04794.

PR XX 26-FEB-1999; 99US-0122152.

PR XX 08-MAR-1999; 99US-0123399.

PR XX 12-JUL-1999; 99US-0332171.

XX XX (CELL-) CELLONICS INC.

XX XX Giuliano KA, Kapur R;

PI XX WPI; 2000-594086/56.

DR XX N-PSDB; AAA93351.

XX XX Automated cell-based characterization of toxin by contacting cells containing luminescent reporter molecules with test substance and analyzing optically

PT XX

PT XX

PT XX

PT XX

PS XX Example 11; Page 173-174; 336pp; English.

XX XX

CC XX The invention relates to systems, methods and reagents for cell-based screening or detection of compounds which affect particular biological functions. The methods of the invention utilise fluorescent biodeceptor molecules which, when acted on by a compound of interest, cause an alteration in the cellular distribution of at least the fluorescent moiety. In one embodiment, the biosensors comprise heat shock proteins (HSPs) fused to a fluorescent protein (e.g., jellyfish green fluorescent protein (GFP), or derivatives thereof). Such biosensors are located in the cytoplasm, but on stress activation translocate to the nucleus. In another embodiment biodeceptor proteins can be used to detect protease activity. Such protease biodeceptor fusion proteins comprise one or more fluorescent proteins; a recognition signal which is cleaved by the protease; and at least one cellular localisation signal. The latter two components may be components of a single protein which is acted upon by the protease, or may be from heterologous sources. Due to the localisation signal, the biodeceptor protein is localised to a particular region of the cell. Once acted on by the protease of interest, the fluorescent protein is cleaved from the localisation sequence, and is free to migrate to other locations within the cell. The presence of a second localisation signal attached to the fluorescent protein enables the fluorescent protein to be directed to a different cellular compartment after cleavage of the protease recognition sequence. The change in distribution of the fluorescent protein can be detected using imaging methods with a high degree of spatial resolution. The methods and biosensors of the invention can be used to investigate a wide range of cellular activities and to screen compounds which modulate these activities. Biosensors containing a recognition site for caspase, for example, may be used for the screening of compounds which modulate apoptosis, while biosensors containing other protease recognition sites may be used for the detection of proteolytic toxins (such as anthrax lethal factor). The method provides improved target validation and candidate compound optimisation by combining many cell screening formats with fluorescence-based molecular reagents and computer-based feature extraction, data analysis and automation, resulting in increased quantity and speed of data collection and faster evaluation of drug candidates. Sequences AAB22860-B22876 and AAB22936-B22941 represent biosensor fusion proteins produced in an exemplification of the invention.

XX XX

XX XX Sequence 294 AA:

Query Match 50.3%; Score 1288.5; DB 21; Length 294;

Best Local Similarity 91.0%; Pred. No. 1.6e-100;

Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

OY 1 MSKGEELFTGVVILVELDGVNGHKFSVSGEGEDATYGLTKLTKFKICTTKLPVPPPTL 60

Db :|||||

2 VSKGEELFTGVVILVELDGVNGHKFSVSGEGEDATYGLTKLTKFKICTTKLPVPPPTL 61

OY 61 VTTLTYGVCFSRYPDEHKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120

Db 62 VTTLTYGVCFSRYPDEHKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121

OY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180

Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD 181

OY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTTAAAGITTHGMDELYKSG 240

Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMHVLLEFVTTAAAGITTHGMDELYKSG 241

OY 241 GSGSGSGSGSGSGG-QSGSGSGSGSG 267

Db 242 LRSAGAGAGAGAGADEVDGAGADEVDG 269

Search completed: June 17, 2003, 16:40:58

Job time : 94.3226 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:29:13 ; Search time 32.5844 Seconds
(without alignments)
430.719 Million cell updates/sec

Title: US-09-863-901-5
 Perfect score: 2560
 Sequence: 1 MSKGEELFTGVVPILVELDG.....LGFVYTMDPACEKLYGGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Issued_Patents_AA:*
1: /cgn2_6/pdata1/1/aa/5A_COMB.pcp:*
2: /cgn2_6/pdata1/1/aa/5B_COMB.pcp:*
3: /cgn2_6/pdata1/1/aa/6A_COMB.pcp:*
4: /cgn2_6/pdata1/1/aa/6B_COMB.pcp:*
5: /cgn2_6/pdata1/1/aa/PCTUS_COMB.pcp:*
6: /cgn2_6/pdata1/1/aa/backfiles1.pcp:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	1290.5	50.4	890	4	US-09-513-783A-174	Sequence 174, App	
2	1288.5	50.3	294	4	US-09-513-783A-2	Sequence 2, Appli	
3	1286	50.2	1070	4	US-09-091-042A-2	Sequence 2, Appli	
4	1283	50.1	459	4	US-09-513-783A-170	Sequence 170, App	
5	1270	49.6	238	3	US-08-893-327-16	Sequence 16, Appl	
6	1270	49.6	247	3	US-08-893-327-18	Sequence 18, Appl	
7	1269	49.6	783	4	US-09-513-783A-176	Sequence 176, App	
8	1267	49.5	941	4	US-09-513-783A-172	Sequence 172, App	
9	1266	49.5	1452	4	US-09-127-227-2	Sequence 2, Appli	
10	1265.5	49.4	350	4	US-09-513-783A-30	Sequence 30, Appl	
11	1265	49.4	805	4	US-09-513-783A-178	Sequence 178, App	
12	1265	49.4	1407	4	US-08-974-549A-628	Sequence 628, App	
13	1263	49.3	302	4	US-09-513-783A-18	Sequence 18, Appl	
14	1262	49.3	238	1	US-08-337-915A-2	Sequence 2, Appli	
15	1262	49.3	238	4	US-09-121-539-1	Sequence 1, Appli	
16	1262	49.3	238	5	PCT-US95-14692-2	Sequence 2, Appli	
17	1259.5	49.2	642	2	US-08-818-253-2	Sequence 2, Appli	
18	1259.5	49.2	642	4	US-08-818-252-2	Sequence 2, Appli	
19	1258.5	49.2	652	4	US-08-818-233-4	Sequence 4, Appli	
20	1259.5	49.2	652	4	US-08-818-252-4	Sequence 4, Appli	
21	1259	49.2	295	4	US-08-818-252-4	Sequence 4, Appli	
22	1258	49.1	238	1	US-09-513-783A-20	Sequence 20, Appl	
23	1258	49.1	238	2	US-08-753-143-2	Sequence 2, Appli	
24	1258	49.1	238	2	US-08-676-865-2	Sequence 2, Appli	
25	1258	49.1	238	2	US-08-680-876-2	Sequence 2, Appli	
26	1258	49.1	238	3	US-08-792-553-2	Sequence 2, Appli	
27	1258	49.1	238	4	US-08-753-144-2	Sequence 2, Appli	
28	1258	49.1	238	4	US-09-094-359-2	Sequence 2, Appli	

ALIGNMENTS

RESULT 1

US-09-513-783A-174

US 03 013 783A 174
: Sequence 174, Application US/09513783A

; Patent No. 6416959

GENERAL INFORMATION:

APPLICANT: **Giuliano, Kenneth A.**

APPLICANT: Kapur, Ravi

; TITLE OF INVENTION: A System for Cell Based Screening

; FILE REFERENCE: 97-022-L1

; CURRENT APPLICATION NUMBER: US/09/513,783A

; CURRENT FILING DATE: 2000-02-25
 ; NUMBER OF SEC ID NOS: 190

```

; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In War 3

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 174

```

```

; SEQ ID NO 174
; LENGTH: 390

```

```

; LENGTH: 390
; TYPE: PR'T

```

TYPE: PRI
ORGANISM: A

ORGANISM: ALLICIAL sequence
; FEATURE:

OTHER IN

US-09-513-783A-174

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match	50.4%	Score 1290.5;	DB 4;	Length 890;
Best Local Similarity	64.4%	Pred. No. 5e-100;		
Matches 282;	Conservative	22;	Mismatches 77;	Indels 57;
Gaps 12;				

[illegible]

Db 353 ---GETKSYFEEVSSMLTKMEIAEAYLGKTVNAVTV--PAYENDSQOATKDAGT 407
QY 385 YAK-----NEPTLIRI 395
Db 408 IAGLVNLRINEPTAAAI 425

RESULT 2

US-09-513-783A-2
; Sequence 2, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: GFP-DEVD-Annexin II construct
US-09-513-783A-2

Query Match 50.3%; Score 1288.5; DB 4; Length 294;
Best Local Similarity 91.0%; Pred. No. 1.6e-100;
Matches 244; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 MSKGEELFGVVPILVELDGVNGHKFVSVEGEGDATYGLTLFICTTGKLPVWPPTL 60
Db 2 VSGEELFGVVPILVELDGVNGHKFVSVEGEGDATYGLTLFICTTGKLPVWPPTL 61
QY 61 VTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNKTRAEVKEGDTLV 120
Db 62 VTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNKTRAEVKEGDTLV 121
QY 121 NRLELKGIDFKEDGNILGHKLEYNSHNHYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180
Db 122 NRLELKGIDFKEDGNILGHKLEYNSHNHYIMADKQNGIKANFKIRHNIEDGVSQVLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDELYKSG 240
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDELYKSG 241
QY 241 GSGSGGSGSGSGG-QSGSGSGGSGG 267
Db 242 LRSGAGAGAGAGAGADEVDG 269

RESULT 3

US-09-091-042A-2
; Sequence 2, Application US/09091042A
; Patent No. 6455300
; GENERAL INFORMATION:

APPLICANT: The Government of the United States of America
as represented by the Secretary
Department of Health and Human Services
Washington, D.C.
Hcun Ph.D., Han
Hager Ph.D., Gordon L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MONITORING
DNA BINDING MOLECULES IN LIVING CELLS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Needle & Rosenberg

STREET: 127 Peachtree Street, Suite 1200

CITY: Atlanta

STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,042A
FILING DATE: 08-Jun-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,373
FILING DATE: 08 Dec 1995
ATTORNEY/AGENT INFORMATION:
NAME: Selby, Elizabeth
REGISTRATION NUMBER: 38298
REFERENCE/DOCKET NUMBER: 14014.0183
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-688-0770
TELEFAX: 404-688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1070 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-091-042A-2

Query Match 50.2%; Score 1286; DB 4; Length 1070;
Best Local Similarity 97.6%; Pred. No. 1.6e-99;
Matches 239; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSKGEELFGVVPILVELDGVNGHKFVSVEGEGDATYGLTLFICTTGKLPVWPPTL 60
Db 23 MSKGEELFGVVPILVELDGVNGHKFVSVEGEGDATYGLTLFICTTGKLPVWPPTL 82
QY 61 VTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNKTRAEVKEGDTLV 120
Db 83 VTLLTYGVQCFSRYPDHMKQHDFFKSAMPEGVQVQERTIFFKDDGNKTRAEVKEGDTLV 142
QY 121 NRLELKGIDFKEDGNILGHKLEYNSHNHYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180
Db 143 NRLELKGIDFKEDGNILGHKLEYNSHNHYIMADKQNGIKANFKIRHNIEDGVSQVLAD 202
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDELYKSG 240
Db 203 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLLFEVTAAGITHGMDELYKSG 262
QY 241 GSGSG 245
Db 263 GAGAG 267

RESULT 4

US-09-513-783A-170
; Sequence 170, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:

APPLICANT: Giuliano, Kenneth A.
APPLICANT: Kapur, Ravi
TITLE OF INVENTION: A System for Cell Based Screening
FILE REFERENCE: 97-022-L1

CURRENT APPLICATION NUMBER: US/09/513,783A

CURRENT FILING DATE: 2000-02-25

NUMBER OF SEQ ID NOS: 180

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 170

LENGTH: 459

TYPE: PRT

ORGANISM: Artificial Sequence

Db 1275 VTTTGYVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 1334
Qy 121 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 180
Db 1335 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 1394
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELK 238
Db 1395 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELK 1452

RESULT 10
US-09-513-783A-30
; Sequence 30, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 30
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: NLS-Fred25-cellubrevin construct
US-09-513-783A-30

Query Match 49.4%; Score 1265.5; DB 4; Length 350;
Best Local Similarity 93.4%; Pred. No. 1.7e-98;
Matches 239; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy 2 SKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 61
Db 9 SKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 68
Qy 62 TLTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLVN 121
Db 69 TLTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLVN 128
Qy 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLADH 181
Db 129 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLADH 188
Qy 182 YQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELKSGG 241
Db 189 YQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELKSGG 248
Qy 242 SGSGGSGSGSGSGSGGS 257
Db 249 S-TGVPSGSSAATGSN 263

RESULT 11
US-09-513-783A-178
; Sequence 178, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 178
; LENGTH: 805

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-NFKB
US-09-513-783A-178

Query Match 49.4%; Score 1265; DB 4; Length 805;
Best Local Similarity 98.8%; Pred. No. 6.1e-98;
Matches 237; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 60
Db 2 VSKGEELFTGVVPILVELDGVNKGHSVSGEGDATYGKLTFLKFICTTGKLPVPMPTLV 61
Qy 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 120
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKFEGDTLV 121
Qy 121 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLEYNHNHYIMADKQKNGIKANFKIRHNTEDEGSVOLAD 181
Qy 181 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELKSG 240
Db 182 HYQONTPIGDGPVLLPDNHYLSTQSALSCKDPNEKRDMHVLLEFVTAAGITGHGMDELKSG 241

RESULT 12
US-08-974-549A-628
; Sequence 628, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Ted
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 628:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1407 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1407
; OTHER INFORMATION: /note= "fusion protein composed of
; OTHER INFORMATION: enhanced green fluorescent protein
; OTHER INFORMATION:
; OTHER INFORMATION: 5' untranslated region of hTERT mRNA and
; OTHER INFORMATION: hTERT protein sequence"
; US-08-974-549A-628

Query Match          49.4%; Score 1265; DB 4; Length 1407;
Best Local Similarity 98.8%; Pred. No. 1.3e-97;
Matches 237; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVPPPTL 60
DB 2 VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVPPPTL 61
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 120
DB 62 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQKNGIKRKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQKNGIKRKIRHNIEDGSVQLAD 181
QY 181 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITGHMDELYKSG 240
DB 182 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITGHMDELYKSG 241

RESULT 13
US-09-513-783A-18
; Sequence 18, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-11
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18

Query Match          49.3%; Score 1263; DB 4; Length 302;
Best Local Similarity 88.0%; Pred. No. 2.3e-98;
Matches 241; Conservative 4; Mismatches 25; Indels 4; Gaps 1;

QY 2 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVPPPTLV 61
DB 3 SKGEELFTGVVPILVELDGDVNGHKFSVSGEGDATYKLTFLKFICTTGKLPVPPPTLV 62
QY 62 TTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLVN 121
DB 63 TTLCYGVQCFSRYPDHMKRHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGDTLVN 122
QY 122 RIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQKNGIKRKIRHNIEDGSVQLADH 181
DB 123 RIELKGIDFKEDGNILGHKLEYNHNVIYIMADKQKNGIKRKIRHNIEDGSVQLADH 182
QY 182 YQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITGHMDELYKSG- 240
DB 183 YQONTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITGHMDELYKSG- 241
QY 241 ---GSGGGGSGGGSGGGSGGGSGGGSGGGSGGS 271
DB 243 RKRKRSAGDEVDAAGDEVDAAGDEVDAAGDEVDAAGS 276

RESULT 14
US-08-337-915A-2
; Sequence 2, Application US/08337915A
; Patent No. 5625048
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Heim, Roger
; TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carson
; STREET: 201 No. 5625048th Figueroa Street, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90012
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,915A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spitals, John P.
; REGISTRATION NUMBER: 29,215
; REFERENCE/DOCKET NUMBER: 1279-178
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-337-915A-2
```


Query Match 49.3%; Score 1262; DB 1; Length 238;
Best Local Similarity 98.7%; Pred. No. 2e-98;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 60

QY 61 VTTLTGVQCFSYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYINADKOKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYINADKOKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238

RESULT 15

US-09-121-539-1
; Sequence 1, Application US/09121539B
; Patent No. 6194548
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshiro
; APPLICANT: Tsukamoto, No. 61945481yo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: 046124-5005
; CURRENT APPLICATION NUMBER: US/09/121,539B
; CURRENT FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: JP 026418/1998
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; OTHER INFORMATION: Green fluorescent protein
US-09-121-539-1

Query Match 49.3%; Score 1262; DB 4; Length 238;
Best Local Similarity 98.7%; Pred. No. 2e-98;
Matches 235; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 60
DB 1 MSKGEELFTGVVPIVLVELDGVNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVWPPTL 60

QY 61 VTTLTGVQCFSYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYINADKOKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVIYINADKOKNGIKANFKIRHNIEDGSVOLAD 180

QY 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDLYK 238

Search completed: June 18, 2003, 17:32:55
Job time : 34.5844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 17:31:48 ; Search time 57.9279 Seconds
(without alignments)
891.015 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2560
Sequence: 1 MSKGEELFTGVVPILVELDQ.....LGFWTMDPACKEKLYGGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1295	50.6	719	9	US-10-072-036-51
2	1292.5	50.5	631	9	US-10-072-036-39
3	1290.5	50.4	890	9	US-10-100-957A-174
4	1288.5	50.3	294	9	US-10-100-957A-2
5	1286	50.2	1070	9	US-10-001-486B-2
6	1286	50.2	1099	9	US-10-259-864-4
7	1286	50.2	1147	9	US-10-259-864-1
8	1283.5	50.1	633	9	US-10-072-036-45
9	1283	50.1	459	9	US-10-100-957A-170
10	1278.5	49.9	635	9	US-10-072-036-125
11	1276.5	49.9	1171	9	US-10-072-036-131
12	1271	49.6	397	9	US-10-033-717-30
13	1271	49.6	429	9	US-10-033-717-29
14	1269.5	49.6	359	9	US-10-033-717-33
15	1269.5	49.6	359	9	US-10-033-717-34
16	1269.5	49.6	391	9	US-10-033-717-32
17	1269.5	49.6	997	9	US-10-072-036-121
18	1269	49.6	607	9	US-10-072-036-47
19	1269	49.6	783	9	US-10-100-957A-176

20	1268	49.5	403	9	US-10-033-717-31	Sequence 31, Appl
21	1268	49.5	842	9	US-10-072-036-43	Sequence 43, Appl
22	1267	49.5	605	9	US-10-072-036-41	Sequence 41, Appl
23	1267	49.5	727	9	US-10-072-036-139	Sequence 139, App
24	1267	49.5	941	9	US-10-100-957A-172	Sequence 172, App
25	1266.5	49.5	806	9	US-10-072-036-53	Sequence 53, Appl
26	1266	49.5	933	9	US-10-072-036-135	Sequence 135, App
27	1266	49.5	1089	9	US-10-259-864-2	Sequence 2, Appl1
28	1266	49.5	1452	12	US-10-050-673-2	Sequence 2, Appl1
29	1265.5	49.4	350	9	US-10-100-957A-30	Sequence 30, Appl1
30	1265	49.4	308	9	US-10-033-717-35	Sequence 35, Appl
31	1265	49.4	442	9	US-10-072-036-127	Sequence 127, App
32	1265	49.4	544	9	US-10-072-036-115	Sequence 115, App
33	1265	49.4	797	9	US-10-072-036-143	Sequence 143, App
34	1265	49.4	805	9	US-10-100-957A-178	Sequence 178, App
35	1265	49.4	843	9	US-10-072-036-117	Sequence 117, App
36	1265	49.4	871	9	US-10-072-036-109	Sequence 109, App
37	1265	49.4	968	9	US-10-072-036-49	Sequence 49, Appl
38	1265	49.4	1039	9	US-10-072-036-55	Sequence 55, Appl
39	1265	49.4	1090	9	US-10-259-864-6	Sequence 6, Appl1
40	1265	49.4	1407	9	US-10-044-592-334	Sequence 334, App
41	1265	49.4	1407	9	US-10-044-539-334	Sequence 334, App
42	1263	49.3	302	9	US-10-100-957A-18	Sequence 18, Appl
43	1262	49.3	238	9	US-09-866-538-2	Sequence 2, Appl1
44	1262	49.3	238	9	US-09-900-345A-125	Sequence 125, App
45	1262	49.3	238	9	US-10-121-258-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-072-036-51
; Sequence 51, Application US/10072036
; Publication NO. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP
; APPLICANT: Sara BJORN
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072.036
; CURRENT FILING DATE: 2002-09-13
; PRIOR FILING DATE: 09/417,197
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-Smad2 fusion
US-10-072-036-51

Query Match	50.6%	Score 1295;	DB 9;	Length 719;
Best Local Similarity	81.0%;	Pred. No. 7.9e-82;		
Matches 251;	Conservative	9;	Mismatches 36;	Indels 14;
Gaps	1;			
Qy	1	MSKGEELFTGVVPILVELDGVNKHKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL	60	
Db	2	VSKGEELFTGVVPILVELDGVNKHKFSVSGEGDATYKGLTLKFTCTTGKLPVWPPTL	61	
Qy	61	VTILTYGVQCFSRYPDHMKQHDFFKSPAMPEGYVQERIFFKDDGNKTRAEVKEGDTLV	120	
Db	62	VTTLYGVQCFSRYPDHMKQHDFFKSPAMPEGYVQERTIFFKDDGNKTRAEVKEGDTLV	121	
Qy	121	NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKANFKIRHNIEDGSVQLAD	180	
Db	122	NRIELKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKANFKIRHNIEDGSVQLAD	181	

Query Match	50.5%	Score 1292.5;	DB 9;	Length 631;
Best Local Similarity	58.0%;	Prd; Mismatches	107;	Indels 71; Gaps 11;
Matches 283;	Conservative	27;		
QY	1	MSKGEELFTGVVPLLYVELGDVNGHKFSVSGEGEDATYGKLTLKFCTTGTGKLPVPWPTL	60	
Db	2	VSKGEELFTGVVPLLYVELGDVNGHKFSVSGEGEDATYGKLTLKFCTTGTGKLPVPWPTL	61	
QY	61	VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFRKDDGNKYTRAEVFEGETLV	120	
Db	62	VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGVYVQERTIFFRKDDGNKYTRAEVFEGETLV	121	
QY	121	NRIELKGIDPKDGNILGHKLEYNNSHNHYIMADKQNKIKANFKIRHNIEDGSVOLAD	180	
Db	122	NRIELAGIDPKDGNILGHKLSYNTNSHNHYIMADKQNKIKVNFIRHNIEDGSVOLAD	181	
QY	181	HYOQNTPIGDGPVLLDPDNHYLSTQSALSKDPNEKRDMHVLEFFTAAGITHGMDELYKSG	240	
Db	182	HYOQNTPIGDGPVLLDPDNHYLSTQSALSKDPNEKRDMHVLEFFTAACITLGMDELYKSG	241	
QY	241	--GGSGGGGGSGGGGGSGGGSGGGSGG-----GQSGGG-----S	280	
Db	242	LRSRAQASNTWMAAAQGSGGGERPRTEGYGPGVEVMVKQPFDVGPRYTQLQYIG	301	
QY	281	GGOSGLRS-----VKLTSDFNPNRWIG-----RHKKMFNFDVNIN	316	
Db	302	EGAYGMVSAYDHVRKTRVAIKSPPEHQYCQTLREIQILLRPFHENVIQIRDLRA	361	
QY	317	GKISLDENVYKASDIVINNLGATPEQAKRHKDAVEAFGG--AGMYKVGETWPAYIEGW	374	
Db	362	STLEARDNYIVODLMETDYLKLLKSQOLSNDHCICYELYQILRGLK-----YIHS	412	

RESULT 4
US-10-100-957A-2
; Sequence 2, Application US/10100957A
; Publication No. US20030096322A1
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.

Db 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
QY 241 GSGSG 245
Db 241 GAGAG 245

RESULT 7

US-10-259-864-1
; Sequence 1, Application US/10259864
; Publication No. US20030077645A1
; GENERAL INFORMATION:
; APPLICANT: Hager, Gordon L
; TITLE OF INVENTION: Superfamily Receptor Chimeras, Translocation Assay For Superfamily
; TITLE OF INVENTION: Receptor Ligands, and Methods and Kits For Detecting and Character
; FILE REFERENCE: 24799
; CURRENT APPLICATION NUMBER: US/10/259,864
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US 60/325,178
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Rat/human chimera
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1)..()
; OTHER INFORMATION: Chimeric Protein
US-10-259-864-1

Query Match 50.2%; Score 1286; DB 9; Length 1147;
Best Local Similarity 97.6%; Pred. No. 5.7e-81;
Matches 239; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVWPPTL 60
Db 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVWPPTL 60
QY 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 121 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
QY 241 GSGSG 245
Db 241 GAGAG 245

RESULT 8

US-10-072-036-45
; Sequence 45, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP

; APPLICANT: Sara BJRON
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072,036
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/417,197
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-Jnk1 fusion
US-10-072-036-45

Query Match 50.1%; Score 1283.5; DB 9; Length 633;
Best Local Similarity 49.6%; Pred. No. 4.3e-81;
Matches 298; Conservative 40; Mismatches 104; Indels 159; Gaps 17;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVWPPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTKLTKFICTTGKLPVWPPTL 61
QY 61 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
QY 121 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 180
Db 122 NRIELKGIDFEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVOLAD 181
QY 181 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 240
Db 182 HYQONTPTGDPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFVTAAGITHGMDELYKSG 239
QY 241 GSG 286
Db 240 -SGLRSARARIMRSRDRNNFYSGEIGDSTFTVLKRYQNLKPIGSGAQGTVCAYDAILE 298
QY 287 RSV---KLTSDFDNPRWIGR-----HKMFNFDVNNHNGKISLDEM--VY---- 326
Db 299 RNAIKLSRPFQNTAKRAYRELVLKMCVNHKNIGLLNVFTPOK-SLEEFQDVIYVM 357
QY 327 -----KASDIVI----- 333
Db 358 ELMANLCQVIQIMELDHERMSYLLYQMLCGIKHLHSAGIHRDLKPSNIVKSDCTLKL 417
QY 334 -----NLLGATPEQAKRKHDVAEAFGGAGMKYGVETDWA----- 369
Db 418 DFLGARTAGTSFMTPIYVTVRYRAPEVL-GMGYKENVDL-WSVGCINGENVCHKILFP 475
QY 370 ---YIEGWKKLATD-----ELEKAYKNP-----TLIRIWDALFDIVDKD 407
Db 476 GRDYIDQWKNVIEQLGTPCPEFMKKLQPTVRYVYNNRPRYAGYSFEKLPDVLFP-ADSE 534
QY 408 QNGAITLDWKAYTAAGIIQSSECEETFRVCDIDESQLDVDE-MTRQHLGFYWTMDP 466
Db 535 HNK-----LKAQARDLLSKMLVIDASKRISVDEALQHPYINWY--DP 576
QY 467 A 467
Db 577 S 577

RESULT 9

US-10-100-957A-170
; Sequence 170, Application US/10100957A

```
; Publication No. US20030096322A1
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-LJA
; CURRENT APPLICATION NUMBER: US/10/100,957A
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 170
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
US-10-100-957A-170

Query Match      50.1%; Score 1283; DB 9; Length 459;
Best Local Similarity 66.8%; Pred. No. 3.2e-81;
Matches 266; Conservative 11; Mismatches 55; Indels 66; Gaps 6;

Qy 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 61

Qy 61 VTTLYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 120
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 121

Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVTIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVTIMADKQNGIKANFKIRHNIEDGSVOLAD 181

Qy 181 HYQONTPIGDGPVLLPNHLYSTQSALSCKDPNEKRDMVLLEFVTAAGITGHMDLYKSG 240
Db 182 HYQONTPIGDGPVLLPNHLYSTQSALSCKDPNEKRDMVLLEFVTAAGITGHMDLYKSG 241

Qy 241 GSGSGGSGSGSGS-----GGQS-----G 258
Db 242 LRSRAASRAESAMTERVPFSLRGPWDPRDWPYHSRFLDQAFLPLRPEWSOWLG 301

Qy 259 GSGSGG-----OSGSGSGSGSGSGSGSLRSVKLTSDFPNPRWGRHKHMF 308
Db 302 GSWPGYVRPLPAAIESPAAPAYSRALSSQSGVSEIRHTAD----RWVS----- 352

Qy 309 NFDLVNNGKISLDEMYKASDIVNNLNGATPEQAKRH 346
Db 353 --LDVNH---PAPDELTVKTKDGVVEITGKHEERQDEH 385

RESULT 10
US-10-072-036-125
; Sequence 125, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP
; APPLICANT: Sara BURON
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An I
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072.036
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/417,197
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 125
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-NFAT fusion
US-10-072-036-131

Query Match      49.9%; Score 1276.5; DB 9; Length 1171;
Best Local Similarity 90.7%; Pred. No. 2.7e-80;
Matches 244; Conservative 3; Mismatches 21; Indels 1; Gaps 1;

Qy 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 61
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-VASP fusion
US-10-072-036-125

Query Match      49.9%; Score 1278.5; DB 9; Length 635;
Best Local Similarity 65.1%; Pred. No. 9.5e-81;
Matches 267; Conservative 23; Mismatches 77; Indels 43; Gaps 7;

Qy 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 61

Qy 61 VTTLYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 120
Db 62 VTTLYGVQCFSRYPDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 121

Qy 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVTIMADKQNGIKANFKIRHNIEDGSVOLAD 180
Db 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVTIMADKQNGIKANFKIRHNIEDGSVOLAD 181

Qy 181 HYQONTPIGDGPVLLPNHLYSTQSALSCKDPNEKRDMVLLEFVTAAGITGHMDLYKSG 240
Db 182 HYQONTPIGDGPVLLPNHLYSTQSALSCKDPNEKRDMVLLEFVTAAGITGHMDLYK-- 239

Qy 241 GSGSGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSLRSVKLTSDFPNPRW 300
Db 240 -----SGLRSRAQASMSSETVIMSETVICSSRATVMLYDDGDKRW 278

Qy 301 I--GRHKHMFLOVNH-----GKISLDEMYKASDIV--INNLCATPEQAKRH 346
Db 279 LPAGTGPQAFSRVQIYHNPTANSFRVGRKKMQPDQVVINCAIVRGVKNQATP-NFHOW 337

Qy 347 KDVAEAFEGGAGKRYGVETDWPAYIEGWKKLATDELEYAKNEPTLIRW 396
Db 338 RDAQVW---GLNFGSKEDAAQAAAGMAS--ALEALEGGGPPPPPALPTW 382
```

RESULT 11

```
US-10-072-036-131
; Sequence 131, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
; APPLICANT: Ole THASTRUP
; APPLICANT: Sara BURON
; APPLICANT: Soren TULLIN
; APPLICANT: Kasper ALMHOLT
; APPLICANT: Kurt SCUDDER
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
; FILE REFERENCE: 3759-0120P
; CURRENT APPLICATION NUMBER: US/10/072.036
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/417,197
; PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 131
; LENGTH: 1171
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-NFAT fusion
US-10-072-036-131
```

```
Query Match      49.9%; Score 1276.5; DB 9; Length 1171;
Best Local Similarity 90.7%; Pred. No. 2.7e-80;
Matches 244; Conservative 3; Mismatches 21; Indels 1; Gaps 1;

Qy 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 60
Db 2 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGEDATYGLTLKFKICTTGKLPVPPWPTL 61
```


APPLICANT: TOPOI, LILIA
APPLICANT: MARX, MARIA
APPLICANT: CALOTHY, GEORGES
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DRM, A SECRETED PROTEIN
TITLE OF INVENTION: WITH CELL GROWTH INHIBITING ACTIVITY
FILE REFERENCE: 14014.0358
CURRENT APPLICATION NUMBER: US/10/033,717
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/444,066
PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/277,407
PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/079,440
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33
LENGTH: 359
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: /No. US20030078406ale -
OTHER INFORMATION: synthetic construct
US-10-033-717-33

Query Match 49.6%; Score 1269.5; DB 9; Length 359;
Best Local Similarity 89.0%; Pred. No. 2e-80;
Matches 242; Conservative 4; Mismatches 7; Indels 19; Gaps 1;
QY 1 MSKGEELFTGVVPILVELDQDVGNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVPWPTL 60
DB 2 VSKGEELFTGVVPILVELDQDVGNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVPWPTL 61
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 62 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITLGMDELYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITLGMDELYKSG 241
QY 241 -----GSGSGGSGSGS 253
DB 242 LRMRAQNDSEQTQSPPPQPSRTRGRGGRGT 273

RESULT 15
US-10-033-717-34
Sequence 34, Application US/10033717
Publication No. US20030078406A1
GENERAL INFORMATION:
APPLICANT: BLAIR, DONALD
APPLICANT: CLAUSEN, PETER
APPLICANT: TOPOI, LILIA
APPLICANT: MARX, MARIA
APPLICANT: CALOTHY, GEORGES
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DRM, A SECRETED PROTEIN
TITLE OF INVENTION: WITH CELL GROWTH INHIBITING ACTIVITY
FILE REFERENCE: 14014.0358
CURRENT APPLICATION NUMBER: US/10/033,717
CURRENT FILING DATE: 2001-12-27
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/444,066
PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/277,407
PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/079,440
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 34
LENGTH: 359
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: /No. US20030078406ale -
OTHER INFORMATION: synthetic construct
US-10-033-717-34
Query Match 49.6%; Score 1269.5; DB 9; Length 359;
Best Local Similarity 89.0%; Pred. No. 2e-80;
Matches 242; Conservative 4; Mismatches 7; Indels 19; Gaps 1;
QY 1 MSKGEELFTGVVPILVELDQDVGNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVPWPTL 60
DB 2 VSKGEELFTGVVPILVELDQDVGNGHKFVSVEGEGDATYKGLTKFKICTTGKLPVPWPTL 61
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 120
DB 62 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKEGDTLV 121
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
DB 122 NRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLAD 181
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITLGMDELYKSG 240
DB 182 HYQONTPIGDGPVLLPDNHYLSTQSALSADPNKRDHMLLEFVTAAGITLGMDELYKSG 241
QY 241 -----GSGSGGSGSGS 253
DB 242 LRMRAQNDSEQTQSPPPQPSRTRGRGGRGT 273

Search completed: June 18, 2003, 17:44:06
Job time : 58.9279 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:38:20 ; Search time 45.2562 Seconds
(without alignments)
1013.257 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2560
Sequence: 1 MSKGELFGVVPILVELDG.....LGFWTMDPACELKLGAVP 477
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.73:**

1: Pirl:**

2: Pirl:**

3: Pirl:**

4: Pirl:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1256	49.1	238	1 JQ1514	green-fluorescent
2	1032	40.3	136	1 AQLFNV	aequorin precursor
3	962	37.6	136	2 A26623	aequorin-1 precurs
4	742	29.0	198	2 S39022	mitrocomin precurs
5	663	25.9	198	2 S28860	claytin - hydromedu
6	170	6.6	302	2 C84470	hypothetical prote
7	168	6.6	592	2 E82759	endo-1,4-beta-gluc
8	161	6.3	115	2 D61615	sericin MG-1 - gre
9	159.5	6.2	207	2 T07381	glycine-rich prote
10	158.5	6.2	268	1 CIHUL	calpain (EC 3.4.22
11	155.5	6.1	1226	2 T24045	hypothetical prote
12	155	6.1	641	1 QQB331	nuclear antigen EB
13	155	6.1	1218	2 E84537	hypothetical prote
14	154.5	6.0	622	2 I37984	keratin 9, type I,
15	154	6.0	1585	2 T31611	hypothetical prote
16	151.5	5.9	481	2 A35628	loricrin - mouse
17	149.5	5.8	204	2 T09592	protein corA, cold
18	149.5	5.8	221	2 T04592	glycine-rich cell
19	148.5	5.8	171	2 H84709	probable glycine-r
20	148	5.8	291	1 S31415	glycine-rich prote
21	148	5.8	1275	2 T49362	hypothetical prote
22	148	5.8	1901	2 F70806	hypothetical glyci
23	147.5	5.8	165	1 KNR2G1	glycine-rich cell
24	147.5	5.8	434	1 Z3BP1K	coat protein A - p
25	147	5.7	183	2 PN0109	keratin-like prote
26	147	5.7	526	1 KRBOVI	keratin, 54K type
27	147	5.7	569	1 KRME51	keratin, 59K type
28	147	5.7	995	2 T22942	hypothetical prote
29	146.5	5.7	266	1 CIRBL	calpain (EC 3.4.22

30	146	5.7	465	1 S01820	glycine-rich cell
31	145.5	5.7	186	1 KRBO2B	keratin, 68K type
32	145	5.7	263	2 A34466	calpain (EC 3.4.22
33	145	5.7	320	2 T09555	fibrillarin - Arab
34	145	5.7	988	2 T08102	myrosinase-binding
35	144.5	5.6	266	1 CIPGL	calpain (EC 3.4.22
36	144.5	5.6	462	4 S33798	FUS/CHOP mutant fu
37	144	5.6	183	1 KNR2G2	glycine-rich cell
38	143.5	5.6	208	2 T46896	merozoite surface
39	143.5	5.6	434	2 S08091	gene III protein -
40	143.5	5.6	593	1 KRHUO	keratin 10, type I
41	143	5.6	150	2 C86224	hypothetical prote
42	143	5.6	420	2 A49442	transcription fact
43	142.5	5.6	167	2 S21359	keratin, type I, c
44	142.5	5.6	271	2 S34666	glycine-rich prote
45	142.5	5.6	528	2 G02127	fus-like protein -

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 23-Mar-2001

C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:92175527; PMID:1347277

A:Accession: JS0692

A:Molecule type: DNA

A:Residues: 1-107, 'S', 109-238 <PRA1>

A:CROSS-references: GB:M62654; NID:g1555662; PIDN:AAA27722.1; PID:g1555663

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99, 'F', 101-140, 'L', 142-218, 'V', 220-238 <PRA2>

A:CROSS-references: GB:M62653; NID:g1555660; PIDN:AAA27721.1; PID:g155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-64; 74-122; 132-151; 154-183; 185-200 <PRA3>

R:Inouye, S.; Tsujii, F.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-156, 'P', 158-171, 'K', 173-238 <INO>

A:CROSS-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13, 'V', 15-24, 'Q', 26-44, 'N', 46-153, 'G', 155-156, 'P', 158-171, 'K', 173-227,

A:CROSS-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24, 'Q', 26-29, 'R', 31-83, 'L', 85-153, 'G', 155-156, 'P', 158-171, 'K', 173-208,

A:CROSS-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011

A:Experimental source: clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A', 2-79, 'R', 8

A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation: X-ray crystallography, 1.9 angstroms
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
 C:Genetics:
 A:Gene: GFP
 A:Introns: 69/3; 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/Cross-link: 5-Imidazolinone (Ser-Gly) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 49.1%; Score 1256; DB 1; Length 238;
 Best Local Similarity 97.5%; Pred. No. 3.7e-73;
 Matches 232; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYKLTIFCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDGYKLTIFCTTGKLPVWPPTL 60
 QY 61 VTTLVGVOCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGFGDILV 120
 DB 61 VTTFSGVOCFSRYPDHMKQHDFFKSAMPEGVQERTIFFKDDGNYKTRAEVKGFGDILV 120
 QY 121 NRIELKIDFKEDGNILGHKLEYNYNHNYIMADKQNGIKANFKIRHNIEDGVSQVLAD 180
 DB 121 NRIELKIDFKEDGNILGHKMEYNYNHNYIMADKQNGIKVNFKIRHNIEDGVSQVLAD 180
 QY 181 HYQONTPIGDGVPVLPDNYHLSQALSQALSKDPNEKRDHMLLEFVTAAGITGMDLYK 238
 DB 181 HYQONTPIGDGVPVLPDNYHLSQALSQALSKDPNEKRDHMLLEFVTAAGITGMDLYK 238

RESULT 2

AQJFNV
 Aequorin precursor - hydromedusa (Aequorea victoria)
 C:Species: Aequorea victoria
 C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 26-May-2000
 C:Accession: A03020
 R:Inouye, S.; Noguchi, M.; Sakaki, Y.; Takagi, Y.; Miyata, T.; Iwanaga, S.; Miyata, T.;
 Proc. Natl. Acad. Sci. U.S.A. 82, 3154-3158, 1985
 A:Title: Cloning and sequence analysis of cDNA for the luminescent protein aequorin.
 A:Reference number: A03020; MUID:85216460; PMID:3858813
 A:Accession: A03020
 A:Molecule type: mRNA
 A:Residues: 1-196 <INO>
 A:Cross-references: GB:M11394; NID:g155658; PIDN:AAA27719.1; PID:g155659; GB:L29571; NID:
 A:Experimental source: clone AQ440
 C:Comment: The precise function of residues 1-7 is not known.
 C:Comment: The authors suggest that there are three calcium-binding sites, and that resi-
 C:Comment: Trace amounts of calcium ion trigger the oxidation of the functional chromoph-
 as a blue light.
 C:Comment: Cysteine residues appear to be critical to the ability of this photoprotein to
 n, and mercaptoethanol is necessary to regenerate aequorin from apoaequorin.
 C:Superfamily: calmodulin; calmodulin repeat homology
 C:Keywords: calcium binding; EF hand; luminescence
 F:8-196/Product: aequorin #status predicted <MAT>
 F:111-143/Domain: calmodulin repeat homology <EF1>
 F:147-179/Domain: calmodulin repeat homology <EF2>
 F:31.33.35.37.42/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted
 F:124,126,128,130,135/Binding site: calcium (Asp, Asn, Ala, Glu) #status predicted
 F:160,162,164,166,171/Binding site: calcium (Asp, Ser, Glu, Asp, Ser, Glu, Asp, Ser, Glu) #status predicted

Query Match 40.3%; Score 1032; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 5.7e-59;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
 DB 7 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 66
 QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 407

DB 67 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 126
 QY 408 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467
 DB 127 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 186
 QY 468 CEKLYGGAVP 477
 DB 187 CEKLYGGAVP 196

RESULT 3

A26623
 Aequorin-1 precursor - hydromedusa (Aequorea victoria)
 C:Species: Aequorea victoria
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-May-2000
 C:Accession: A26623
 R:Prasher, D.C.; McCann, R.O.; Longiaru, M.; Cormier, M.J.
 Biochemistry 26, 1326-1332, 1987
 A:Title: Sequence comparisons of complementary DNAs encoding aequorin isotypes.
 A:Reference number: A26623; MUID:87185437; PMID:2882777
 A:Accession: A26623
 A:Molecule type: mRNA
 A:Residues: 1-196 <PRA>
 A:Cross-references: GB:M16103; NID:g155652; PIDN:AAA27716.1; PID:g155653
 A:Note: The authors translated the codon GAT for residue 143 as Ala
 C:Comment: The precise function of residue 1-7 is not known.
 C:Superfamily: calmodulin; calmodulin repeat homology
 C:Keywords: calcium binding; EF hand; luminescence
 F:18-50/Domain: calmodulin repeat homology <EF1>
 F:111-143/Domain: calmodulin repeat homology <EF2>
 F:147-179/Domain: calmodulin repeat homology <EF3>

Query Match 37.6%; Score 962; DB 2; Length 196;
 Best Local Similarity 90.5%; Pred. No. 1.7e-54;
 Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 347
 DB 7 SVKLTSDFDNPRWIGRHKHMFNLDVNHNGKISLDEMYKASDIVINNIGATPEQAKRHK 66
 QY 348 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 407
 DB 67 DAVEAFPGGAGMYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRINGDALFDIVDKD 126
 QY 408 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467
 DB 127 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 186
 QY 468 CEKLYGGAVP 477
 DB 187 CEKLYGGAVP 196

RESULT 4

S39022
 Mitrocomin precursor - Mitrocoma cellularia
 C:Species: Mitrocoma cellularia
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-May-2000
 C:Accession: S39022
 R:Fagan, T.F.; Ohmiya, Y.; Blinks, J.R.; Inouye, S.; Tsujii, F.I.
 FEBS Lett. 333, 301-305, 1993
 A:Title: Cloning, expression and sequence analysis of cDNA for the Ca(2+)-binding pro-
 A:Reference number: S39022; MUID:94039781; PMID:8224198
 A:Accession: S39022
 A:Molecule type: mRNA
 A:Residues: 1-198 <PAG>
 A:Cross-references: EMBL:L31623; NID:9468907; PIDN:AAA29298.1; PID:g468908
 C:Superfamily: calmodulin; calmodulin repeat homology
 C:Keywords: calcium binding; EF hand; luminescence
 F:19-51/Domain: calmodulin repeat homology <EF1>
 F:112-144/Domain: calmodulin repeat homology <EF2>
 F:148-180/Domain: calmodulin repeat homology <EF3>

Query Match 29.0%; Score 742; DB 2; Length 198;
Best Local Similarity 67.9%; Pred. No. 1.8e-40;
Matches 129; Conservative 29; Mismatches 32; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFLDVNVHGKISLDEMVKASDITVINNLGATPEQAKRHK 347
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 8 AVKLTTDFDNPKWIARHKHMFNFDINSNQINLNEMVHKASNIIICKLGATEEQTRHQ 67

QY 348 DAVEAFEGGAGMKYGVEWDMPAYIEGWKKLATDELEKYAKNEPTLIIRWGDALEDIVDKD 407
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 68 KCVEDFEGGAGLEYDKDTTWPEYIEGWKKLAETELERHSKNQVTILRIWGDALDIIDKD 127

QY 408 ONCAITLDEWKATKAAGIIOSSDECETEFRVCDDIDESGOLDVDVDMTROHLGFYWTMDPA 467
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 128 RNGSVSLEDEWQTTHCAGIQSRGQCCEATFAHCDDGGDKLDVDMTROHLGFYSVDPT 187

QY 468 CEKLYGGAVP 477
|||:|||||:
DB 188 CEGLYGGAVP 197

RESULT 5
S28860
clytin - hydromedusa (*Clytia gregarium*)
C:Species: *Clytia gregarium*
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S28860
R:inouye, S.; Tsuji, F.I.
FEBS Lett. 315, 343-346, 1993
A:Title: Cloning and sequence analysis of cDNA for the Ca(2+)-activated photopor-
A:Reference number: S28860; MUID:93138101; PMID:8422928
A:Accession: S28860
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-198 <INO>
A:Cross-references: GB:I13247; NID:g469233; PIDN:AAA28293.1; PID:g469234
C:Superfamily: calmodulin; calmodulin repeat homology
C:Keywords: EF hand; luminescence
F:20-52/Domain: calmodulin repeat homology <EF1>
F:113-145/Domain: calmodulin repeat homology <EF2>
F:149-181/Domain: calmodulin repeat homology <EF3>

Query Match 25.9%; Score 663; DB 2; Length 198;
Best Local Similarity 61.6%; Pred. No. 2e-35;
Matches 117; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFLDVNVHGKISLDEMVKASDITVINNLGATPEQAKRHK 347
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 AVKLRPENFDNPKVNRHKHMFNFNDINGDGKITLDEIVSKASDDICAKLGATPEQTRKHQ 68

QY 348 DAVEAFEGGAGMKYGVEWDMPAYIEGWKKLATDELEKYAKNEPTLIIRWGDALEDIVDKD 407
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 69 DAVEAFFKIGMDYGYGEVEFPAPVDVGKKLANLDLKWSQNKKSLRDWGEAEVDFDKD 128

QY 408 ONCAITLDEWKATKAAGIIOSSDECETEFRVCDDIDESGOLDVDVDMTROHLGFYWTMDPA 467
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 129 GSGSISLDEWKAYGRISGCSSDEDAEKTFKHCDLNSGKLDVDMTROHLGFYTLDPN 188

QY 468 CEKLYGGAVP 477
|||:|||||:
DB 189 ADGLYNFVP 198

RESULT 6
C84470
hypothetical protein At2g05580 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 01-Mar-2002
C:Accession: C84470
R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; T-
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.;

Cross-references: GDB:119752; OMIM:114170

244 SGGSGGSGGGQSSGSGGGSGG 282

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:31:10 ; Search time 23.5332 Seconds
(without alignments)
840.693 Million cell updates/sec

Title: US-09-863-901-5

Perfect score: 2560

Sequence: 1 MSKGELFTGVPIVLVDG.....LGFWTMDPACELKYGAVP 477

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1262	49.3	238	1	GFP_AEQVI
2	1032	40.3	196	1	AE02_AEQVI
3	962	37.6	196	1	AE01_AEQVI
4	742	29.0	198	1	MYTR_MITCE
5	718	28.0	195	1	OBL_OBELO
6	663	25.9	198	1	CLYT_CLYGR
7	171.5	6.7	532	1	ZIC2_HUMAN
8	161.5	6.3	450	1	SWPI_ENCCU
9	161	6.3	115	1	SER1_GALME
10	158.5	6.2	268	1	CANS_HUMAN
11	155	6.1	641	1	EBN1_EBV
12	154.5	6.0	622	1	K1C1_HUMAN
13	151.5	5.9	481	1	LORI_MOUSE
14	149.5	5.8	204	1	CORA_MEDSA
15	149	5.8	269	1	CANS_MOUSE
16	149	5.8	421	1	BR3A_MOUSE
17	148.5	5.8	627	1	K2C1_MOUSE
18	148	5.8	1380	1	DDX9_MOUSE
19	148	5.8	1901	1	Y208_MYCTU
20	147.5	5.8	165	1	GRP1_ORYSA
21	147.5	5.8	434	1	CORA_BPIKE
22	147	5.7	526	1	K1C1_BOVIN
23	147	5.7	569	1	K1C1_MOUSE
24	146.5	5.7	266	1	CANS_RABIT
25	146	5.7	465	1	GRP2_PHAVU
26	145.5	5.7	166	1	K2C5_BOVIN
27	145	5.7	263	1	CANS_BOVIN
28	144.5	5.6	266	1	CANS_PIG
29	144	5.6	183	1	GRP2_ORYSA
30	143.5	5.6	266	1	CANS_RAT
31	143.5	5.6	434	1	COAA_BPI22
32	143.5	5.6	593	1	K1CJ_HUMAN
33	143.5	5.6	677	1	SP87_DICDI

RESULT 1

ID	GFP_AEQVI	STANDARD;	PRT;	238 AA.
AC	P42212; Q17104;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Green fluorescent protein.			
GN	GFP.			
OS	Aequorea victoria (Jellyfish).			
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;			
OC	Aequoreidae; Aequorea.			
OX	NCBI_TaxID=6100;			
RP	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=92175527; PubMed=1347277;			
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,			
RA	Cormier M.J.;			
RT	"Primary structure of the Aequorea victoria green-fluorescent			
RT	protein.";			
RL	Gene 111:229-233(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94185810; PubMed=8137953;			
RA	Inouye S., Tsuji F.I.;			
RT	"Aequorea green fluorescent protein. Expression of the gene and			
RT	fluorescence characteristics of the recombinant protein.";			
RL	FEBS Lett. 341:277-280(1994).			
RN	[3]			
RP	CHROMOPHORE.			
RX	MEDLINE=93192221; PubMed=8448132;			
RA	Cody C.W., Frasher D.C., Westler W.M., Prendergast F.G., Ward W.W.;			
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea			
RL	green-fluorescent protein.";			
RN	Biochemistry 32:1212-1218(1993).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=96355665; PubMed=8703075;			
RA	Ormo M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,			
RA	Remington S.J.;			
RT	"Crystal structure of the Aequorea victoria green fluorescent			
RT	protein.";			
RL	Science 273:1392-1395(1996).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98294543; PubMed=9631087;			
RA	Yang F., Moss L.G., Phillips G.N. Jr.;			
RT	"The molecular structure of green fluorescent protein.";			
RL	Nat. Biotechnol. 14:1246-1251(1996).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.			
RX	MEDLINE=98455509; PubMed=9782051;			
RA	Wachter R.M., Eisliger M.A., Kallio K., Hanson G.T., Remington S.J.;			
RT	"Structural basis of spectral shifts in the yellow-emission variants			
RT	of green fluorescent protein.";			
RL	Structure 6:1267-1277(1998).			

ALIGNMENTS

[7]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE-99238303; PubMed-10220315;
 RA Elslinger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 RT "Structural and spectral response of green fluorescent protein
 RT variants to changes in pH.";
 RL Biochemistry 38:5296-5301(1999).
 CC -!- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
 CC BLUE CHEMILUMINESCENCE OF THE PROTEIN Aequorin INTO GREEN
 CC FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
 CC RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN Aequorin.
 CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
 CC ABSORBANCE PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
 CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
 CC -!- SUBUNIT: MONOMER.
 CC -!- TISSUE SPECIFICITY: PHOTOCYTES.
 CC -!- PTM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
 CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
 CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
 CC -!- BIOTECHNOLOGY: Has become a useful and ubiquitous tool for making
 CC chimeric proteins of GFP linked to other proteins where it
 CC functions as a fluorescent protein tag. GFP tolerates N- and C-
 CC terminal fusion to a broad variety of proteins. It has been
 CC expressed in bacteria, yeast, slime mold, plants, drosophila,
 CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
 CC marker in living cells, it allows for a wide range of applications
 CC where it may function as a cell lineage tracer, reporter of gene
 CC expression, or as a measure of protein-protein interactions.
 CC -!- DATABASE: NAME-Protein Spotlight;
 CC NOTE-Issue 11 of June 2001;
 CC WWW="http://www.expasy.org/spotlight/articles/split011.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M62654; AAA27722.1; -;
 DR EMBL; M62653; AAA27721.1; -;
 DR EMBL; L29345; AAA58246.1; -;
 DR PDB; 1GFL; 11-JAN-97.
 DR PDB; 1EMA; 08-NOV-96.
 DR PDB; 1EMB; 16-JUN-97.
 DR PDB; 1EMC; 20-AUG-97.
 DR PDB; 2EMD; 20-AUG-97.
 DR PDB; 1EME; 20-AUG-97.
 DR PDB; 1EMF; 20-AUG-97.
 DR PDB; 1EMG; 12-MAY-99.
 DR PDB; 1EMK; 20-AUG-97.
 DR PDB; 1EML; 20-AUG-97.
 DR PDB; 1EMN; 20-AUG-97.
 DR PDB; 2EMO; 20-AUG-97.
 DR PDB; 1BFP; 07-JUL-97.
 DR PDB; 1YFP; 28-OCT-98.
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR PRINTS; PR01229; GFP; 1.
 DR PRODOM; PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT SITE 65 67
 FT VARIANT 100 100
 FT VARIANT 108 108
 FT VARIANT 141 141
 FT VARIANT 219 219
 FT VARIANT 25 25
 FT CONFLICT 157 157
 FT CONFLICT 172 172
 FT CONFLICT 238 AA; 26886 MW; EA5AGF21FBFB6E05 CRC64;
 SQ SEQUENCE

Query Match 49.3%; Score 1262; DB 1; Length 238;
 Best Local Similarity 98.7%; Pred. No. 3e-69;
 Matches 235; Conservative 1; Mismatches 0; Gaps 0;
 QY 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGSDATYKLTLLKFTCTTGKLPVPMPTL 60
 DB 1 MSKGEELFTGVVPILVELDGVNCHKFSVSGEGSDATYKLTLLKFTCTTGKLPVPMPTL 60
 QY 61 VTTLTGYVQCFSPRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDFLV 120
 DB 61 VTTTSYGVCFSRYPDHMKQDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKEGDFLV 120
 QY 121 NRIELKGIDFEDKNILGHKLEYNNSHYVIMADKQKNGIKRHNIEDGSVQLAD 180
 DB 121 NRIELKGIDFEDKNILGHKLEYNNSHYVIMADKQKNGIKRHNIEDGSVQLAD 180
 QY 181 HYQONTPTIGDGPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITGMDELYK 238
 DB 181 HYQONTPTIGDGPVLLPDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITGMDELYK 238
 RESULT 2
 AEQ2_AEQV1 STANDARD; PRT; 196 AA.
 ID AEQ2_AEQV1
 AC P02592;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aequorin 2 precursor.
 OS Aequorea victoria (Jellyfish).
 CC Eukaryota; Metazoa; Chndaria; Hydrozoa; Leptomedusae;
 CC Aequoreidae; Aequorea.
 CC NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85215460; PubMed-3858813;
 RA Inouye S., Noguchi M., Sakaki Y., Takagi Y., Miyata T., Iwanaga S.,
 RA Miyata T., Tsuji F.I.;
 RT "Cloning and sequence analysis of cDNA for the luminescent protein
 RT aequorin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3154-3158(1985).
 RN [2]
 RP SEQUENCE OF 9-185 FROM N.A. (AEQUORIN 2 AND 3).
 RX MEDLINE-87185437; PubMed-2882777;
 RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;
 RT "Sequence comparisons of complementary DNAs encoding aequorin
 RT isoforms.";
 RL Biochemistry 26:1326-1332(1987).
 RN [3]
 RP SEQUENCE OF 8-196.
 RX MEDLINE-86077721; PubMed-2866797;
 RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,
 RA Cormier M.J., Vanaman T.C.;
 RT "Amino acid sequence of the calcium-dependent photoprotein aequorin.";
 RL Biochemistry 24:6762-6771(1985).
 RN [4]
 RP MUTAGENESIS.
 RA Tsuji F.I., Inouye S., Goto T., Sakaki Y.;
 RT "Site-specific mutagenesis of the calcium-binding photoprotein
 RT aequorin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8107-8111(1986).
 RN [5]
 RP MUTAGENESIS OF PRO-196.
 RX MEDLINE-92111761; PubMed-1765170;
 RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;
 RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-
 RT binding photoprotein, aequorin.";
 RL FEBS Lett. 295:63-66(1991).
 RN [6]
 RP DISULFIDE BOND.
 RX MEDLINE-94009705; PubMed-8405461;
 RA Ohmiya Y., Kurono S., Ohashi M., Fagan T.F., Tsuji F.I.;
 RT "Mass spectrometric evidence for a disulfide bond in aequorin

regeneration.";
FEBS Lett. 332:226-228(1993).
[7]
X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
MEDLINE-20289815; PubMed-10830969;
Head J.F., Inouye S., Teranishi K., Shimomura O.;
"The crystal structure of the photoprotein aequorin at 2.3-A
resolution.";
Nature 405:372-376(2000).
-1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
COELENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE
CONCOMITANT EMISSION OF LIGHT.
-1- PTM: THE REDUCTION OF THE DISULFIDE BOND IS NECESSARY TO
REGENERATE AEQUORIN FROM APOAEQUORIN.
-1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
-1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L29571; AAA27720.1; -
EMBL; M16104; AAA27717.1; -
EMBL; M16105; AAA27718.1; -
EMBL; M11394; AAA27719.1; -
PIR; A03020; AQFNV.
PDB; 1EJ3; 31-MAY-00.
InterPro; IPR002048; EF-hand.
Pfam; PF00036; efhand; 3.
ProDom; PD000012; EF-hand; 1.
SMART; SM00054; Eph; 2.
PROSITE; PS00018; EF_HAND; 3.
Photoprotein; Calcium-binding; Luminescence; Repeat; 3D-structure.
FT CHAIN 1 7
FT SITE 8 196
FT SITE 47 57
FT SITE 62 72
FT SITE 107 117
FT CA_BIND 31 42
FT DOMAIN 72 88
FT CA_BIND 124 135
FT CA_BIND 160 171
FT DISULFID 152 159
FT SITE 196 196
FT VARIANT 70 71
FT VARIANT 164 164
FT MUTAGEN 36 36
FT MUTAGEN 129 129
FT MUTAGEN 165 165
FT MUTAGEN 65 65
FT MUTAGEN 152 152
FT MUTAGEN 152 152
FT MUTAGEN 159 159
FT MUTAGEN 187 187
FT CONFLICT 37 37
SQ SEQUENCE 196 AA; 22285 MW; 532DC7A9D29BA80C CRC64;
Query Match 40.3%; Score 1032; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.5e-55;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 288 SVLTSDFDNPRIGRKHMFNLDVNHNGKISLDENVYKASDIVINLNGATPEQAKRHK 347
|||||
D 7 SVLTSDFDNPRIGRKHMFNLDVNHNGKISLDENVYKASDIVINLNGATPEQAKRHK 66
QY 348 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRWGDAFDIVDKD 407
|||||

Db 67 DAVEAFGGAGMYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRWGDAFDIVDKD 126
QY 408 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGOLDVDEMTROHLGFYWTMDPA 467
|||||
Db 127 QNGAITLDEWKAYTKAAGIIQSSDECEETFRVCDIDESGOLDVDEMTROHLGFYWTMDPA 186
QY 468 CEKLYGGAVP 477
Db 187 CEKLYGGAVP 196
|||||
RESULT 3
AEQ1_AEQVI STANDARD; PRT; 196 AA.
AC P07164;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aequorin 1 precursor.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87185437; PubMed-2882777;
RA Prasher D.C., McCann R.O., Longiaru M., Cormier M.J.;
RT "Sequence comparisons of complementary DNAs encoding aequorin
RT isotypes.";
RL Biochemistry 26:1326-1332(1987).
RN [2]
RP SEQUENCE OF 8-196.
RX MEDLINE-8607721; PubMed-2866797;
RA Charbonneau H., Walsh K.A., McCann R.O., Prendergast F.G.,
RA Cormier M.J., Vanaman T.C.;
RT "Amino acid sequence of the calcium-dependent photoprotein aequorin.";
RL Biochemistry 24:6762-6771(1985).
RN [3]
RP MUTAGENESIS OF PRO-196.
RX MEDLINE-92111761; PubMed-1765170;
RA Nomura M., Inouye S., Ohmiya Y., Tsuji F.I.;
RT "A C-terminal proline is required for bioluminescence of the Ca(2+)-
RT binding photoprotein, aequorin.";
RL FEBS Lett. 295:63-66(1991).
RN [4]
RP DISULFIDE BOND.
RX MEDLINE-94009705; PubMed-8405461;
RA Ohmiya Y., Kurono S., Ohashi M., Fagan T.F., Tsuji F.I.;
RT "Mass spectrometric evidence for a disulfide bond in aequorin
RT regeneration.";
RL FEBS Lett. 332:226-228(1993).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COELENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- PTM: THE REDUCTION OF THE DISULFIDE BOND IS NECESSARY TO
CC REGENERATE AEQUORIN FROM APOAEQUORIN.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M16103; AAA27716.1; -
PIR; A26623; A26623.
DR HSSP; P02592; 1EJ3.
DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00036; ehand; 3.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT FT PROPEP 1 7
FT CHAIN 8 196
FT SITE 47 57
FT SITE 62 72
FT SITE 107 117
FT CA-BIND 31 42
FT DOMAIN 72 88
FT CA-BIND 124 135
FT CA-BIND 160 171
FT DSULFID 152 159
FT SEQUENCE 196 AA; 22514 MW; 9AA5B636288A5B8F CRC64;

Query Match 37.6%; Score 962; DB 1; Length 196;
Best Local Similarity 90.5%; Pred. No. 2.3e-51;
Matches 172; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHK 347
Db 7 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHK 66

QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 407
Db 67 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 126

QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467
Db 127 QNGAISLDEWKAYTKSDGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 186

QY 468 CEKLYGGAVP 477
Db 187 CEKLYGGAVP 196

RESULT 4
MYTR_MITCE STANDARD; PRT; 198 AA.
ID MYTR_MITCE STANDARD; PRT; 198 AA.
AC P39047;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitrocomin precursor.
GN M17.
OS Mitrocoma cellularia (Halistaura mitrocoma).
OC Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroids; Anthomedusae;
OC Mitrocomidae; Mitrocoma.
OX NCBI_TaxID=31874;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94039781; PubMed=82241198;
RA Fagan T.F., Ohmura Y., Blinks J.R., Inouye S., Tsuji F.I.;
RT Cloning, expression and sequence analysis of cDNA for the Ca(2+)-
binding photoprotein, mitrocomin.";
RL FEBS Lett. 333:301-305(1993).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
COLENTHERAZINE INTO COLENTHERAMIDE AND CO(2) WITH THE
CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin family.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; L31623; AAA29298.1; .
DR PIR; S39022; S39022.
DR PIR; P02592; IEJ3.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 3.
DR ProDom; PD000012; EF-hand; 1.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT FT PROPEP 1 8
FT CHAIN 9 198
FT CA-BIND 32 43
FT DOMAIN 73 89
FT CA-BIND 125 136
FT CA-BIND 161 172
FT DSULFID 153 160
FT SEQUENCE 198 AA; 22714 MW; 8F6307EF0966F670 CRC64;

Query Match 29.0%; Score 742; DB 1; Length 198;
Best Local Similarity 67.9%; Pred. No. 3.6e-38;
Matches 129; Conservative 29; Mismatches 32; Indels 0; Gaps 0;

QY 288 SVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHK 347
Db 8 AVKLTDFDNPWGRHKHMFNFVDVNHNGKISLDEMVKASDIVINNIGATPEQAKRHK 67

QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 407
Db 68 KCVDFGGAGLEKDKTTPPEYIEGWKKLATDELEKAKNEPTLIRIWGDALFDIVDKD 127

QY 408 QNGAITLDEWKAYTKAAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 467
Db 128 RGVSLDEWKAYTKAGIIQSSDCEETFRVCDIDESGQLDVDEMTROHLGFWYTMDDPA 187

QY 468 CEKLYGGAVP 477
Db 188 CEKLYGGAVP 197

RESULT 5
OBL_OBELO STANDARD; PRT; 195 AA.
ID OBL_OBELO STANDARD; PRT; 195 AA.
AC Q27709;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Obelin precursor (OBL).
OS Obelia longissima (Black sea hydrosoma).
OC Eukaryota; Metazoa; Chordata; Hydrozoa; Hydroids; Leptomedusae;
OC Campanulariidae; Obelia.
OX NCBI_TaxID=32570;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95180731; PubMed=7875600;
RA Illarionov B.A., Bondar V.S., Illarionova V.A., Vysotski E.S.;
RT "Sequence of the cDNA encoding the Ca(2+)-activated photoprotein
Obelin from the hydroid polyp Obelia longissima.";
RL Gene 153:273-274(1995).
CC -1- FUNCTION: CA(++)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
COLENTHERAZINE INTO COLENTHERAMIDE AND CO(2) WITH THE
CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. Aequorin family.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

or send an email to license@lsb-sib.ch.

```

CC EMBL: U07128; AAA67708.1; -.
CC HSSP: P02592; 1EJ3.
DR InterPro: IPR002048; EF-hand.
DR PFam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 6 POTENTIAL.
FT CHAIN 7 195 OBELIN.
FT CA_BIND 30 41 EF-HAND 1 (POTENTIAL).
FT DOMAIN 71 87 ANCESTRAL CALCIUM SITE 2.
FT CA_BIND 123 134 EF-HAND 3 (POTENTIAL).
FT CA_BIND 159 170 EF-HAND 4 (POTENTIAL).
FT DISULFID 151 158 BY SIMILARITY.
SQ SEQUENCE 195 AA; 22226 MW; 5D002270B73D3663 CRC64;

Query Match 28.0%; Score 718; DB 1; Length 195;
Best Local Similarity 68.4%; Pred. No. 9.8e-37;
Matches 130; Conservative 23; Mismatches 37; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFDVNHNGKISLDEWYKASDIVINNLCATPEQAKRHK 347
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 6 AVKLTDFDNPRWIKRHKHMFDFLDINGNGKITLDEIVSKASDDICAKLEATPEQTKRHQ 65
QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 66 VCVEAFPGCGMEYGEIAFPQFLDQWKLATSEKLNWNEPTLIRENGDAVDFIDFKD 125
QY 408 QNGAITLDEWKATKAAGIIQSSDCEETFRVCDIDESQGLDYDENTROHLGFWYTMDDPA 467
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 126 GSGTITLDEWKAYKRGISGSPSQEDCEATFRHCDLNSGLDYDENTROHLGFWYTLDP 185
QY 468 CEKLYGAVP 477
DB :|||:|||||:
DB 186 ADGLYGNVP 195

```

```

RESULT 6
CLYT_CLYGR STANDARD; PRT; 198 AA.
AC Q08121;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clytin precursor (Phalidin).
OS Clytia gregaria (Phalidium gregarium).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Leptomedusae;
OC Campanulariidae; Clytia.
OX NCBI_TaxID=27801;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93138101; PubMed=8422928;
RA Inouye S., Tsuji F.I.;
RT Cloning and sequence analysis of cDNA for the Ca(2+)-activated
RT photoprotein, clytin.;
RL FEBS Lett. 315:343-346(1993).
CC -1- FUNCTION: CA(2+)-DEPENDENT BIOLUMINESCENCE PHOTOPROTEIN. DISPLAYS
CC AN EMISSION PEAK AT 470 NM (BLUE LIGHT). TRACE AMOUNTS OF CALCIUM
CC ION TRIGGER THE INTRAMOLECULAR OXIDATION OF THE CHROMOPHORE,
CC COELENTERAZINE INTO COELENTERAMIDE AND CO(2) WITH THE
CC CONCOMITANT EMISSION OF LIGHT.
CC -1- SIMILARITY: BELONGS TO THE EF-HAND SUPERFAMILY. AEQUORIN FAMILY.
CC -1- SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/)
CC

```

or send an email to license@lsb-sib.ch.

```

CC EMBL: L13247; AAA28293.1; -.
CC DR EMBL: X70221; CAA49754.1; -.
DR PIR: S28860; S28860.
DR HSSP: P02592; 1EJ3.
DR InterPro: IPR002048; EF-hand.
DR PFam: PF00036; ehand; 3.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF-HAND; 3.
KW Photoprotein; Calcium-binding; Luminescence; Repeat.
FT PROPEP 1 9 POTENTIAL.
FT CHAIN 10 198 CLYTIN.
FT CA_BIND 33 44 EF-HAND 1 (POTENTIAL).
FT DOMAIN 74 90 ANCESTRAL CALCIUM SITE 2.
FT CA_BIND 126 137 EF-HAND 3 (POTENTIAL).
FT CA_BIND 162 173 EF-HAND 4 (POTENTIAL).
FT DISULFID 148 161 BY SIMILARITY.
SQ SEQUENCE 198 AA; 22541 MW; 23F1E399667F9059 CRC64;

Query Match 25.9%; Score 663; DB 1; Length 198;
Best Local Similarity 61.6%; Pred. No. 2e-33;
Matches 117; Conservative 36; Mismatches 37; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNFDVNHNGKISLDEWYKASDIVINNLCATPEQAKRHK 347
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 AVKLTDFDNPRWIKRHKHMFDFLDINGNGKITLDEIVSKASDDICAKLEATPEQTKRHQ 68
QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 69 DAVEAFKKIGMDYIGREVEFPFADGWKELATDLKLSQNKSLINDWGEAVDFIDFKD 128
QY 408 QNGAITLDEWKATKAAGIIQSSDCEETFRVCDIDESQGLDYDENTROHLGFWYTMDDPA 467
DB :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 129 GSGSILDEWKAYGRISGICSDDEAEKTFKHDLSNGKLDYDENTROHLGFWYTLDPN 188
QY 468 CEKLYGAVP 477
DB :|||:|||||:
DB 189 ADGLYGNVP 198

```

```

RESULT 7
ZIC2_HUMAN STANDARD; PRT; 532 AA.
AC Q95409; Q9H309;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein ZIC2 (zinc finger protein of the cerebellum 2).
GN ZIC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT HPES POLY-ALA INSERTION.
RX MEDLINE=98442655; PubMed=9771712;
RA Brown S.A., Wazburton D., Brown L.Y., Yu C.Y., Roeder E.R.,
RA Stengel-Rutkowski S., Hennekam R.C., Muenke M.;
RT "Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila
RT odd-paired.";
RL Nat. Genet. 20:180-183(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20556339; PubMed=10984499;
RA Yang Y., Hwang C.K., Junn E., Lee G., Mouradian M.M.;
RT "ZIC2 and Sp1 repress Sp1-induced activation of the human D1A dopamine
RT receptor gene.";
RL J. Biol. Chem. 275:38863-38869(2000).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: DEFECTS IN ZIC2 ARE A CAUSE OF HOLOPROSENCEPHALY TYPE 5

```

```

CC (HPE5). HPE5 IS A STRUCTURAL ANOMALY OF THE BRAIN
CC -1- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF104902; AAC96325.1; -
CC EMBL: AF193855; AAG28409.1; -
CC HSP: P08047; ISP2.
CC -----
CC Genew; HGNC:12873; ZIC2.
CC MIM: 603073; -
CC InterPro: IPR000822; Znf_C2H2.
CC Pfam: PF00096; zf-C2H2; 5.
CC ProDom: PD000003; Znf_C2H2; 1.
CC SMART: SM00355; Znf_C2H2; 4.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; 4.
CC Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Disease mutation; Holoprosencephaly.
FT DOMAIN 20 23 POLY-HIS.
FT DOMAIN 25 33 POLY-ALA.
FT DOMAIN 89 97 POLY-ALA.
FT DOMAIN 226 230 POLY-ALA.
FT DOMAIN 231 239 POLY-HIS.
FT DOMAIN 300 415 ZINC FINGERS.
FT ZN_FING 300 327 C2H2-TYPE (ATYPICAL).
FT ZN_FING 333 357 C2H2-TYPE.
FT ZN_FING 363 387 C2H2-TYPE.
FT ZN_FING 393 415 C2H2-TYPE.
FT DOMAIN 456 470 POLY-ALA.
FT DOMAIN 490 508 POLY-GLY.
FT VARIANT A -> AAAAAAAAAA (IN HPE5).
FT /FTID=VAR_008856.
FT CONFLICT 124 128 RFGD -> ARLPT (IN REF. 1).
SQ SEQUENCE 532 AA; 55006 MW; BA3E6455DAF97EAC CRC64;

Query Match 6.7%; Score 171.5; DB 1; Length 532;
Best Local Similarity 31.0%; Pred. No. 0.0017;
Matches 54; Conservative 24; Mismatches 73; Indels 23; Gaps 8;

QY 124 ELKIDFKEDGNILGHKLEYNYNSHNVYI--MADK---QKNGIKANFKIRINIEDGSVQ- 177
DB 366 EFEGCDREFANSRDKKHMVHTSDKPYLCMKCDKSYTHPSSLRKHKMKVHSSPQGSSE 425
QY 178 --LADHYQQNTPIGDGPVLLPDNHYLSTQSLSKDPNEKRDHVMVLEFVTAAGITHGMD 235
DB 426 PAASGYSESPPG---LVSPS---AEPQSSNLSP-----AAAAAAAAAAAAAAVSA 473
QY 236 LKSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 288
DB 474 VRGGSGSGSGAGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 525

RESULT 8
SWP1_ENCUCU
ID SWP1_ENCUCU STANDARD; PRT; 450 AA.
AC Q9XZV1.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Spore wall protein 1 precursor.
GN SWP1.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Bonne W., Ferguson D.J.P., Kohler K., Gross U.;
RT "Molecular characterisation of a developmentally expressed spore wall
RL protein from the human microsporidian Encephalitozoon cuniculi.";
Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SPORE WALL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ133745; CAB39735.1; -
CC Sporulation; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 450 SPORE WALL PROTEIN 1.
FT DOMAIN 66 74 POLY-SER.
FT DOMAIN 79 82 POLY-ARG.
FT DOMAIN 303 339 THR-RICH.
FT DOMAIN 355 450 GLY/SER-RICH.
SQ SEQUENCE 450 AA; 45873 MW; 5E7071A3E3A6DF60 CRC64;

Query Match 6.3%; Score 161.5; DB 1; Length 450;
Best Local Similarity 55.4%; Pred. No. 0.0057;
Matches 41; Conservative 1; Mismatches 5; Indels 27; Gaps 4;

QY 239 SGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 274
DB 373 SDGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 432
QY 275 SG---SGSGSGSGSG 285
DB 433 SGGSGSGSGSGSGSG 446

RESULT 9
SERI_GALME
ID SERI_GALME STANDARD; PRT; 115 AA.
AC O96614;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Sericin-1 (Silk gum protein 1) (Fragment).
GN SERI OR SER-1.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Pyraloidea; Pyralidae; Galleriinae; Galleria.
OX NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Middle silk gland;
RA Zurovec M., Sehgal F., Scheller K., Kumaran A.K.;
RT "Silk gland specific cDNAs from Galleria mellonella L.";
Insect Biochem. Mol. Biol. 22:55-67(1992).
CC -1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC SECTION OF SILK GLANDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF095241; AAC79078.1; -

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M34398; AAA39444.1; -
DR EMBL; U09189; AAA82152.1; -
DR PIR; A35628; A35628.
DR HSP; P02876; 9WCA.
DR MGD; MGI:96816; Lor.
KW Keratinization.
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

Query Match 5.9%; Score 151.5; DB 1; Length 481;
Best Local Similarity 45.2%; Pred. No. 0.024; 33; Indels 11; Gaps 3;
Matches 38; Conservative 2; Mismatches 3;
QY 228 GITHGMDLYKSGSG-SGGQSGSGG-----SGGQSGSGG-----SGGQSGSGGSGGSGGSGG 276
DB 395 GSSGCGGYSGGGCGGSGGSGGCGGCGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 454
QY 277 GSGGSGGSLRSVKLTSDNDPRW 300
DB 455 GGGSGGGRGVPVCHOTQOQAPTW 478

RESULT 14
CORAMEDSA STANDARD; PRT; 204 AA.
AC Q07202;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Cold and drought-regulated protein CORA.
GN CORA.
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Apica;
RX MEDLINE=94143496; PubMed=8310076;
RA Laberge S., Castonguay V., Vezina L.-P.;
RT "New cold- and drought-regulated gene from Medicago sativa.";
RL Plant Physiol. 101:1411-1412(1993)
CC -!- FUNCTION: MAY BE INVOLVED IN RESISTANCE OF THE PLANT TO ENVIRONMENTAL STRESS.
CC -!- INDUCTION: BY COLD, ABSCISIC ACID (ABA) AND DROUGHT STRESS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; L03708; AAA9833.1; -
KW Multigene family; Repeat.
FT DOMAIN 54 176 7 X 6 AA REPEATS OF Y-N-H-G-G-G.
FT REPEAT 54 59 1-1.
FT REPEAT 65 70 1-2.
FT REPEAT 71 76 1-3.
FT REPEAT 78 83 1-4.
FT REPEAT 85 90 1-5.
FT REPEAT 164 169 1-6.
FT REPEAT 171 176 1-7.
FT DOMAIN 98 192 11 X 3 AA REPEATS OF H-G-G.
FT REPEAT 98 100 2-1.
FT REPEAT 101 103 2-2.
FT REPEAT 112 114 2-3.
FT REPEAT 115 117 2-4.
FT REPEAT 126 128 2-5.

FT REPEAT 129 131 2-6.
FT REPEAT 178 180 2-7.
FT REPEAT 181 183 2-8.
FT REPEAT 184 186 2-9.
FT REPEAT 187 189 2-10.
FT REPEAT 190 192 2-11.
SQ SEQUENCE 204 AA; 19599 MW; 8A2C082359FCC17F CRC64;
Query Match 5.8%; Score 149.5; DB 1; Length 204;
Best Local Similarity 35.4%; Pred. No. 0.012;
Matches 35; Conservative 11; Mismatches 44; Indels 9; Gaps 3;
QY 228 GITHGMDLYKSGSG-SGGQSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 283
DB 70 GYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGGYNHGGG 129
QY 284 SGLRSVKL-----TSDFDNPRIWIRHMKHMFNLDVNHNG 317
DB 130 GGAESVAVQTEKTNVNDYAKYGGGNYNDGRGGYNHGG 168

RESULT 15
CANS_MOUSE STANDARD; PRT; 269 AA.
ID CANS_MOUSE STANDARD; PRT; 269 AA.
AC O88456; OSVEK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
DE (Calcium-activated neutral proteinase) (CANP).
GN CAPN1 OR CAPN4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20285452; PubMed=10825211;
RA Arthur J.S.C., Elce J.S., Hegadorn C., Williams K., Greer P.A.;
RT "Disruption of the murine calpain small subunit gene, Capn4: calpain is essential for embryonic development but not for cell growth and division."
RL Mol. Cell. Biol. 20:4474-4481(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding (By similarity).
CC -!- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; AF058298; AAC97194.1; -
DR EMBL; BC018352; AAH18352.1; -
DR HSP; P04632; IKF0.
DR MGD; MGI:88266; Capn1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 3.
DR ProDom; PD000012; EF-hand; 1.
DR PROSITE; PS00018; EF_HAND; 2.
KW Calcium-binding; Repeat.

Search completed: June 17, 2003, 16:41:32
Job time : 25.5332 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2003, 16:36:34 ; Search time 93.2277 Seconds
(without alignments)
1054.241 Million cell updates/sec

Title: US-09-863-901-5
Perfect score: 2560
Sequence: 1 MSGEELFTGVVPILVELDG.....LGFWTMDPACELKYGAVP 477

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1256	49.1	238	5	Q93125
2	1254	49.0	238	5	Q27903
3	1231	48.1	758	4	Q96J01
4	1220	47.7	238	5	Q17105
5	1205	47.1	238	5	Q17106
6	1100	43.0	238	5	Q8WTC6
7	1096	42.8	238	5	Q8WTC5
8	1093	42.7	238	5	Q8WTC4
9	1091	42.6	238	5	Q8WTC3
10	1090	42.6	238	5	Q8WTC9
11	1090	42.6	238	5	Q8WTC8
12	1087	42.5	238	5	Q8WTC7
13	1086	42.4	238	5	Q8WTC5
14	915	35.7	195	5	Q8WY8
15	899	35.1	195	5	Q8WY7
16	705	27.5	195	5	Q8T620

17	253.5	9.9	225	5	Q95UA7
18	246	9.6	225	5	Q963F5
19	240	9.4	236	5	Q8T6U0
20	233.5	9.1	266	5	Q9U6Y3
21	231.5	9.0	225	5	Q8T5F1
22	210	8.2	227	5	Q962P9
23	208	8.1	229	5	Q9U6Y6
24	207.5	8.1	234	5	Q8T5F2
25	206.5	8.1	235	5	Q8T5F0
26	205.5	8.0	232	5	Q9GPI5
27	204.5	8.0	225	5	Q9U6Y8
28	204	8.0	227	5	Q95V70
29	203.5	7.9	238	5	Q9BLT9
30	201.5	7.9	232	5	Q9G2Z8
31	201	7.9	221	5	Q95P04
32	198.5	7.8	225	5	Q8T6T9
33	194	7.6	232	5	Q9U6Y7
34	192	7.5	230	5	Q9GTJ7
35	137.5	7.3	227	5	Q95W85
36	136.5	7.3	228	5	Q9GPI6
37	132.5	7.1	238	5	Q9BLZ0
38	131.5	7.1	233	5	Q963I9
39	179.5	7.0	227	5	Q95W86
40	179.5	7.0	231	5	Q9U6Y5
41	175	6.8	229	5	Q8T5E7
42	173.5	6.8	227	5	Q95W11
43	170	6.6	302	10	Q9SL09
44	170	6.6	738	5	Q02402
45	169	6.6	125	5	Q964C1

ALIGNMENTS

RESULT 1

Q93125 PRELIMINARY; PRT; 238 AA.
AC Q93125;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Green fluorescent protein mutant 3.
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96305137; PubMed=8707053;
RA Cormack B.P., Valdivia R.H., Falkow S.;
RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
RL Gene 173:33-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Cormack B.P., Bertram G., Egerton M., Gow N.A.R., Falkow S.,
RA Brown A.J.P.;
RT "Yeast Enhanced Green Fluorescent Protein (YEGFP): a reporter of gene
expression in Candida albicans.";
RL Microbiology 0:0-0(1996).
DR EMBL; U73901; AAB18957.1; -.
DR HSSP; P42212; 1BFP.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26840 MW; A28622809A9DEA60 CRC64;

Query Match 49.1%; Score 1256; DB:5; Length 238;

Best Local Similarity 98.3%; Pred. No. 6e-82;

Matches 234; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MSGEELFTGVVPILVELDGVNGHKFSVSGEGSDATYGKLT/LKLTCTTCKLPVWPPTL 60

```
|||||
Db 1 MSGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 61 VTTFGVGVQCFARYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYK 238

RESULT 2
Q27903 PRELIMINARY; PRT; 238 AA.
AC Q27903;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
SEQUENCE FROM N.A.
RA Rowland G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent
protein by modification of its codon usage."
RL Plant Mol. Biol. 33:989-999(1997).
DR EMBL; X96418; CAAG5278.1; -.
DR HSSP; P42212; 1GFL.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFPLORESENT.
DR PRODOM; PD013756; Green_fl_protein; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match 49.0%; Score 1254; DB 5; Length 238;
Best Local Similarity 97.9%; Pred. NO. 8.3e-82;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 60
Db 1 MSGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 60
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 61 VTTFGVGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
Db 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYK 238

RESULT 3
Q96JQ1 PRELIMINARY; PRT; 758 AA.
AC Q96JQ1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Green fluorescent protein.
GN Raichu404X.
RA Raichu404X.
```

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RX MEDLINE-21322811; PubMed-11429608;
RA Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA Miyawaki A., Matsuda M.;
RT "Spacio-temporal Images of Growth Factor-induced Activation of Ras and
RT Rapi.";
RL Nature 411:1065-1068(2001).
DR EMBL; AB051846; BAB61868.1; -.
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR001806; Ras_trnsmfmg.
DR InterPro; IPR003116; RBD.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF01353; GFP; 2.
DR Pfam; PF00071; Ras; 1.
DR Pfam; PF02196; RBD; 1.
DR PRODOM; PD013756; Green_fl_protein; 2.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW GTP-binding.
KW GTP-binding.
SQ SEQUENCE 758 AA; 85015 MW; 8612408F607CFD49 CRC64;

Query Match 48.1%; Score 1231; DB 4; Length 758;
Best Local Similarity 94.3%; Pred. NO. 1.9e-79;
Matches 233; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 60
Db 499 VSKGEELFTGVVPIILVELDGVNGHKFSVSGEGDATYKGLTKLFICTTGKLPVPWPTL 558
QY 61 VTTLTGVQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 120
Db 559 VTTLTGWQCFSRYPDHMKQHDFFKSAMPEGVYQERTIFFKDDGNYKTRAEVKFEGLTLV 618
QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 180
Db 619 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQKNGIKANFKIRHNIEDGSVQLAD 678
QY 181 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYKSG 240
Db 679 HYQONTPIGDGPVLLPDNHYLSTQSALS KDPNKRHDHMLVLEFVTAAGITHGMDELYKSG 738
QY 241 GSGSGGQ 247
Db 739 RMSKDGK 745

RESULT 4
Q17105 PRELIMINARY; PRT; 238 AA.
AC Q17105;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
SEQUENCE FROM N.A.
RA Watkins J.N., Campbell A.K.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X83959; CAA58789.1; -.
DR HSSP; P42212; 1GFL.
DR InterPro; IPR000786; Green_fl_protein.
```

```
DR Pfam: PF01353; GFP; 1.
DR PRINTS: PR01229; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
FT NON_TER 238
SQ SEQUENCE 238 AA; 26E2BE450E748E44 CRC64;

Query Match 47.7%; Score 1220; DB 5; Length 238;
Best Local Similarity 94.5%; Pred. No. 2.2e-79;
Matches 223; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
QY 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 5
Q17106 PRELIMINARY; PRT; 238 AA.
ID Q17106
AC Q17106
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein (Fragment).
GN GFP.
OS Aequorea victoria (Jellyfish).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=6100;
RN [1]
RP STRAIN=GFPX19UV;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT *Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.*;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435431; AAL33916.1; -
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26E2BE450E748E44 CRC64;

Query Match 43.0%; Score 1100; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 8.1e-71;
Matches 197; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
QY 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 7
Q8WP95 PRELIMINARY; PRT; 238 AA.
ID Q8WP95
AC Q8WP95
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPX.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN=GFPX, AND GFPXMAX;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013824; AAK02062.1; -
DR EMBL: AY013821; AAK02059.1; -
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
```

```
RESULT 6
Q8WTC6 PRELIMINARY; PRT; 238 AA.
ID Q8WTC6
AC Q8WTC6
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN=GFPX19UV;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
Li S.J., Xia N.S.;
RT *Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.*;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435431; AAL33916.1; -
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26E2BE450E748E44 CRC64;

Query Match 43.0%; Score 1100; DB 5; Length 238;
Best Local Similarity 82.8%; Pred. No. 8.1e-71;
Matches 197; Conservative 20; Mismatches 21; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
DB 1 MSKGEELFTGVVPILVELDGVNQHFKFSVSGEGDGYKGLTKLFCITGKLPVPWPTL 60
QY 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
DB 61 VTTLTYGVQCFSRYPDHMKQHDFFKSAPEGYVOERTIFFKDDGNYKTRAEVFEGLTV 120
QY 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDPKEDGNILGHKLEYNHSHNYINADKQKNGIKANFKIRHNIEDGSVOLAD 180
QY 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238
DB 181 HYQONTPIGDPVLLPDNHYLSQTSALSQSDPNKRDHMLLEFVTAAGITHGMDELYK 238

RESULT 7
Q8WP95 PRELIMINARY; PRT; 238 AA.
ID Q8WP95
AC Q8WP95
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFPX.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RP STRAIN=GFPX, AND GFPXMAX;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY013824; AAK02062.1; -
DR EMBL: AY013821; AAK02059.1; -
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
```

SQ SEQUENCE 238 AA; 27049 MW; 8185D0E5E29012B CRC64;

Query Match 42.8%; Score 1096; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 1.6e-70;
 Matches 196; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 QY 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
 DB 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 8

Q8WTC4 PRELIMINARY; PRT; 238 AA.

ID AC Q8WTC4
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein.
 GN GFP.
 OS Aequorea macrodactyla.
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=147615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GFPM161;
 RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
 RA Li S.J., Xia N.S.;
 RT "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF435427; AAL33912.1;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 42.7%; Score 1093; DB 5; Length 238;
 Best Local Similarity 83.6%; Pred. No. 2.6e-70;
 Matches 199; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 QY 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
 DB 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 9

Q8WTD0 PRELIMINARY; PRT; 238 AA.

ID AC Q8WTD0
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein.
 GN GFP.
 OS Aequorea macrodactyla.
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=147615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GFPM161;
 RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
 RA Li S.J., Xia N.S.;
 RT "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF435427; AAL33912.1;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 238 AA; 26956 MW; 75521EAF0CEBA73A CRC64;

Query Match 42.6%; Score 1091; DB 5; Length 238;
 Best Local Similarity 82.4%; Pred. No. 3.6e-70;
 Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

QY 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 DB 1 MSKGEELFTGVVPIVLVDGVDNGHFKFSVSGEGSDATYKLTLCFCTTGKLPVWPPTL 60
 QY 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 DB 61 VTTLTGYQCFSRYPDHMKQHDFFKSAMPEGYQERTIFFKDDGNKTRAEVKEGDTLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHNYIMADKQNGIKANFKIRHNIEDGSVOLAD 180
 QY 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238
 DB 181 HYQNTPIGDGVPVLLPDNHYLSQTALSADPNKRDHMLVLEFVTAAGITHGMDELYK 238

RESULT 10

Q8WTC9 PRELIMINARY; PRT; 238 AA.

ID AC Q8WTC9
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein.
 GN GFP.
 OS Aequorea macrodactyla.
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=147615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GFPM161;
 RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
 RA Li S.J., Xia N.S.;
 RT "Colorful mutants of green fluorescent protein from Aequorea
 macrodactyla";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF435428; AAL33913.1;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP; 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 238 AA; 27031 MW; 5F80A19C19DC584D CRC64;


```
Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

Qy 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

RESULT 11
Q8WTC8 PRELIMINARY; PRT; 238 AA.
AC Q8WTC8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

Qy 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

RESULT 12
Q8WTC7 PRELIMINARY; PRT; 238 AA.
AC Q8WTC7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

Query Match 42.6%; Score 1090; DB 5; Length 238;
Best Local Similarity 82.4%; Pred. No. 4.2e-70;
Matches 196; Conservative 19; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60
Db 1 MSKGEELFTGVPIVLVDGVDVNGHKFSVSGEGDGYATGKTLKFKICTTGKLPVPWPTL 60

Qy 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTITLYGVQCFSRYPDHMKQHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEVKEGDTLV 120

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDFKEDGNILGHKLEYNNSHNYIMADKQNGKIFKANKIRHNIEDGSVOLAD 180

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

Qy 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238
Db 181 HYQONTPIGDPVLLPDNHYLSQTSALSQKDPNEKRDMVLEFFVTAAGITHGMDELYK 238

RESULT 13
Q8WTC5 PRELIMINARY; PRT; 238 AA.
AC Q8WTC5;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;
OC Aequoreidae; Aequorea.
OX NCBI_TaxID=147615;
RN [1]
RC STRAIN=GFPX1910V;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521EA5534E573A CRC64;

Query Match 42.4%; Score 1086; DB 5; Length 238;
Best Local Similarity 83.2%; Pred. No. 8.1e-70;
```

Matches 198; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

QY 1 MSKGBELTGVVPIILVELDGVNKHGKFSVSGEGDATYKLTLLKFKTCTGKLPVWPPTL 60
DB 1 MSKGBELTGVVPIILVELDGVNKHGKFSVSGEGDADYKLEIKFKTCTGKLPVWPPTL 60
QY 61 VTTLYGVOQSRVSDHMKQHDFFKSAPEGYVQERTIFFKDDGNYKTRAEVKPEGDTLV 120
DB 61 VTTLYGILCFARPEHMKMDFKSAPEGYQERTIFFQDDGKYKTRGEVKEGDTLV 120
QY 121 NRLEKLGIDFKEDGNILGHKLEYNNSHNVYIMADKOKNGIKANFKTRHNIEDSGVQIAD 180
DB 121 NRLEKGMDFKEDGNILGHKLEYNNSHNVYIMPDKANGLKVNFKIRHNIEGGVQIAD 180
QY 181 HYQNTPLDGPVLLPDNHYLSTOSALSKDPNEKRDHMLVLEFVTAAGITHGMDELYK 238
DB 181 HTQTNPLDGPVLLPDNHYLSYQTAISKDRNEYRDHMLVLEFFSACGHTGMDELYK 238

RESULT 14

Q8WQY8 PRELIMINARY; PRT; 195 AA.

AC Q8WQY8;

DT 01-MAR-2002 (TremBLrel. 20, Created)

DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)

DE Apoeaquorin.

GN AEQUORIN.

OS Aequorea parva.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=148610;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AEQXM;

RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL; AY013822; AAK02060.1; -

DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00036; EF-hand; 3.

DR ProDom; PD000012; EF-hand; 1.

DR SMART; SM00054; EFH; 3.

DR PROSITE; PS00018; EF_HAND; UNKNOWN_3.

SQ SEQUENCE 195 AA; 22248 MW; 73BE72E7E0C140BE CRC64;

Query Match 35.7%; Score 915; DB 5; Length 195;
Best Local Similarity 85.3%; Pred. No. 9.7e-58;
Matches 162; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 347
DB 6 AVKLEPDFENPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 65
QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB 66 EAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 125
QY 408 QNGAITLDWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMCPA 467
DB 126 QNGAITLDWKYKTSAGIIQSAEDCEETFKVCDLDDSGRLDADENTRQHIGFWYTMCPA 185
QY 468 CEKLYGGAVP 477
DB 186 CEKLYGGAVP 195

Search completed: June 17, 2003, 16:43:24
Job time : 94.2277 secs

Query Match 35.7%; Score 915; DB 5; Length 195;
Best Local Similarity 85.3%; Pred. No. 9.7e-58;
Matches 162; Conservative 18; Mismatches 10; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 347
DB 6 AVKLEPDFENPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 65
QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB 66 EAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 125
QY 408 QNGAITLDWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMCPA 467
DB 126 QNGAITLDWKYKTSAGIIQSAEDCEETFKVCDLDDSGRLDADENTRQHIGFWYTMCPA 185
QY 468 CEKLYGGAVP 477
DB 186 CEKLYGGAVP 195

RESULT 15

Q8WQY7 PRELIMINARY; PRT; 195 AA.

AC Q8WQY7;

DT 01-MAR-2002 (TremBLrel. 20, Created)

DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)

DE Apoeaquorin.

GN AEQUORIN.

OS Aequorea macrodactyla.

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Leptomedusae;

OC Aequoreidae; Aequorea.

OX NCBI_TaxID=147615;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AEQXM;

RA Luo W.X., Zhang J., Yang H.J., Li S.W., Xie X.Y., Qin Y.X., Pang S.Q.,
Li S.J., Xia N.S.;
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RL EMBL; AY013823; AAK02061.1; -

DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00036; EF-hand; 3.

DR ProDom; PD000012; EF-hand; 1.

DR SMART; SM00054; EFH; 2.

DR PROSITE; PS00018; EF_HAND; UNKNOWN_3.

SQ SEQUENCE 195 AA; 22266 MW; A459391B4B8003BF CRC64;

Query Match 35.1%; Score 899; DB 5; Length 195;
Best Local Similarity 83.7%; Pred. No. 1.4e-56;
Matches 159; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 288 SVKLTSDFNPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 347
DB 6 AVKLEPDFENPRWIGRHKHMFNLDVNHNGKISLDMVYKASDIVINNIGATPEQAKRHK 65
QY 348 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 407
DB 66 DAVEAFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRINGDALFDIVDKD 125
QY 408 QNGAITLDWKAYTKAAGIIQSSDCEETFRVCDIDESGOLDVDEMTROHLGFWYTMCPA 467
DB 126 QNGAITLDWKYKTSAGIIQSAEDCEETFKVCDLDDSGRLDADENTRQHIGFWYTMCPA 185
QY 468 CEKLYGGAVP 477
DB 186 CEKLYGGAVP 195

Search completed: June 17, 2003, 16:43:24
Job time : 94.2277 secs